EPIGENETIC BASIS OF STEM CELL IDENTITY IN NORMAL AND MALIGNANT HEMATOPOIETIC DEVELOPMENT

by
Namyoung Jung

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

July, 2015

Abstract

Acute myeloid leukemia (AML) is a heterogeneous hematologic malignancy characterized by subpopulations of leukemia-initiating or leukemia stem cells (LSC) that give rise to clonally related non-stem leukemic blasts. The LSC model proposes that since LSC and their blast progeny are clonally related, their functional properties must be due to epigenetic differences. In addition, the cell of origin of LSC among normal hematopoietic stem and progenitor cells (HSPCs) has yet to be clearly demonstrated. In order to investigate the role of epigenetics in LSC function and hematopoietic development, we profiled DNA methylation and gene expression of CD34+CD38-, CD34+CD38+ and CD34- cells from 15 AML patients, along with 6 well-defined HSPC populations from 5 normal bone marrows using Illumina Infinium HumanMethylation450 BeadChip and Affymetrix Human Genome U133 Plus 2.0 Array. To define LSC and blast functionally, we performed engraftment assays on the three subpopulations from 15 AML patients and defined 20 LSCs and 24 blast samples. We identified the key functional LSC epigenetic signature able to distinguish LSC from blasts that consisted of 84 differential methylations regions (DMRs) in 70 genes that correlated with differential gene expression. *HOXA* cluster genes were enriched within the LSC epigenetic signature. We found that most of these DMRs involve epigenetic alteration independent of underlying mutations, although several are downstream targets of genetic mutation in epigenome modifying enzymes and upstream regulators. The LSC epigenetic signature could predict overall survival for AML patients independent of known risk factors such as age and cytogenetics. We characterized epigenetic changes during normal human hematopoietic development and identified key novel regulators for

hematopoietic differentiation such as *HMHB1* and *MIR539*. We found that global hypomethylation is a critical mechanism of lineage commitment in human hematopoiesis. Our DNA methylation analysis in human hematopoiesis revealed variable epigenetic regulation compared to murine hematopoiesis. Furthermore, we found that LSC populations formed two distinct clusters resembling either lymphoid-primed multipotent progenitors (L-MPPs) or granulocyte/macrophage progenitors (GMPs). These results provide the first evidence for epigenetic variation between LSC and their blast progeny in AML, and its prognostic power. We also provided comprehensive methylome map of human hematopoiesis and identified epigenetically distinct subgroups of AML LSCs that likely reflect the cell of origin.

Readers:

Ravindra Majeti, M.D., Ph.D.

Andrew P. Feinberg, M.D., M.P.H.

Thesis Advisor:

Andrew P. Feinberg, M.D., M.P.H.

Acknowledgments

Over the past 6 years of my graduate study, many people have contributed to my personal and professional growth. It would be impossible to mention everyone who helped me to finish this long journey in graduate school, but I wish to acknowledge couple people here.

Foremost, I would like to express my deepest gratitude to my thesis advisor, Dr. Andrew Feinberg for his relentless support, patience, and motivation. Andy has given insightful suggestions and comments on any subject that we had a conversion on. He also taught me how I should behave in academic setting as a professional scientist. Without Andy's guidance and persistent help throughout my graduate study, this thesis would not have been possible.

I would like to thank to Dr. Ravindra Majeti at Stanford who has been a wonderful thesis committee member, the second reader of my thesis, and collaborator. Ravi has provided his expert advices and valuable knowledge in leukemia and hematopoiesis throughout this thesis. His comments and suggestions were an enormous help for me to learn knowledge in a different field.

I am deeply grateful to my thesis committee chair, Dr. Roger Reeves for his generous support and encouragement for this thesis and my career, and thesis committee member, Dr. Donald Small for his advices and support.

I wish to acknowledge Dr. Yunje Cho, my undergraduate thesis advisor, for his encouragements and advices, which have been motivation for my PhD work.

I am deeply grateful to my collaborators, Dr. Andrew Gentles and Dr. Rafael Irizarry for their statistical advices and comments, and Dr. Bo Dai for the essential part of functional experiments of this thesis. I would like to express my gratitude toward the AML patients and their family members for their decisions to provide valuable samples for this thesis.

I could not have done my PhD work without the advices and supports provided by current and former Feinberg lab members who have had impact on every aspect of this thesis. Akiko Doi and Brian Herb have always offered great feedbacks for experiments and encouragements to move forward during my graduate course work. It has been very lucky to have Amy Vandiver as my bench mate for past 5 years, who has provided useful suggestions for science, and warm support for any decision related to graduate course work and even personal life. I'm grateful to Carolina Montano, Lindsay Rizzardi, Hwajin Lee, Xin Li, Yun Liu, Hong Ji, Peter Murakami, Michael Multhaup, Varenka Rodriguez, and Elisabet Pujadas for useful discussions, feedbacks, and their friendship. I thank to Rakel Tryggvadottir and Arni Runarsson for their experimental help.

I am grateful to Samsung Scholarship that has offered financial support for first five years, and Mogam Science Scholarship Foundation for financial support for last year of my graduate study.

Many friends have supported and helped me to stay centered throughout past 6 years. I have been fortunate to have warmhearted CMM classmates, who have helped me to adapt to a new culture. Friends of Korean community and in Korea have been always there to listen to my problems, and given me great emotional support.

Last but not least, this long journey in graduate school would have been impossible without the warm support of my family. My grandparents, Jongsung Jung and Oksoon Cho have offered priceless life lessons and taught me how to be a polite person. My parents, Insoo Cheong and Geumsook Choi, have provided unconditional love and support for any decision that I have made. My two younger sisters, Gayoung Jung and Chaeyoung Jung have been the best friends throughout my life giving me emotional support.

I will always be grateful to all the individuals mentioned above who shaped me as a professional scientist.

Table of Contents

List of Tables

List of Figures

Chapter1

Introduction

1. Epigenetics

Overview

Epigenetics is the study of information that is heritable after cell division other than the primary DNA sequence. Epigenetics is involved in many cellular processes such as embryonic development, differentiation, and an interaction with environmental stimuli. Mechanisms of epigenetics include DNA methylation, histone modification, chromatin factors, chromatin structure, and noncoding RNAs. Dysregulation of the epigenetic modifications is known to be involved in many diseases such as cancer.

Definition (historical and modern)

Conrad Waddington first coined the term 'epigenetics' in 1942 as a mechanism of how genotype brought about phenotype during development (Waddington, 2012). He proposed the 'Epigenetic landscape' as a model for cellular development that a ball is rolling down a hill to the lowest points (Goldberg et al., 2007). The 'Epigenetic landscape' is a metaphor of how a cell (a ball) decides its fate (the lowest points) (Goldberg et al., 2007). The role of epigenetics is evident in multicellular organisms, where the vast majority of cells have the same genomic sequence, yet exhibit distinct cellular phenotypes. All cells in the human body originate from the same source, pluripotent stem cells in the inner cell mass (ICM) of a blastocyst. The pluripotent stem cell differentiates into different types of cells such as neurons, hepatocytes, or tubular epithelial cells, which constitute brain, liver, and kidney, respectively. As all these cells come from the pluripotent stem cells in the ICM, they have the same genomic sequence,

yet display distinct phenotypes and functionality (Figure 1.1). In order to investigate the role of epigenetics in different levels of biological processes systematically, NIH Roadmap Epigenomics Project was launched in 2008 (Bernstein et al., 2010). The NIH Roadmap Epigenomics Project defines epigenetics as both heritable and other stable and long-term changes in gene expression regulation of a cell. Collective efforts to provide a public resource of epigenomic information have generated comprehensive epigenomic maps of DNA methylation, histone modifications, DNA binding protein, chromatin accessibility, and noncoding RNAs in diverse cell types and tissues (Bernstein et al., 2010; Roadmap Epigenomics et al., 2015). Among the epigenetic modifications, DNA methylation will be the focus of this thesis.

DNA methylation

In mammals, DNA methylation is generally referred to as a covalent modification in which a methyl group is attached to the fifth position of cytosine in CpG dinucleotides. Note that non-CG methylation in CHG and CHH (H=A, C or T) contexts has been reported in stem cell, neuron, and other differentiated tissues, yet its functional role is still an active research area (Lister et al., 2013; Lister et al., 2009; Ramsahoye et al., 2000; Schultz et al., 2015). DNA methylation is established and maintained by three DNA methyltransferases: DNMT1, DNMT3A and DNMT3B (Li et al., 1992; Okano et al., 1999). These enzymes use S-adenosyl methionine (SAM) as a methyl group donor. DNMT1 is responsible for methylation of hemimethylated DNA after DNA replication, while DNMT3A and DNMT3B are able to methylate both hemi and unmethylated DNA,

so serve as *de novo* methyl transferases (Leonhardt et al., 1992; Okano et al., 1999). The human genome contains about 28 million CpGs and 60-80% of those are generally methylated (Smith and Meissner, 2013). A small portion (~10%) of CpGs, not methylated clusters together, and thus establishes a genomic region called *CpG islands*, which are predominantly located in the promoter of coding genes (Bird, 1986; Deaton and Bird, 2011). Maintaining the unmethylated status of CpG islands requires transcription factor binding, deposition of histone variant H2A.Z, and trimethylation of histone H3 at lysine4 (H3K4me3) that inhibit the binding of DNA methyltransferases (Brandeis et al., 1994; Conerly et al., 2010; Macleod et al., 1994; Otani et al., 2009). The repression of CpG island-promoters genes is largely mediated by histone modification, particularly trimethylation of histone H3 at lysine27 (H3K27me3), and polycomb proteins (Bartke et al., 2010; Brinkman et al., 2012; Jones, 2012). However, DNA methylation at CpG island promoters does occur to suppress CpG island containing genes that are targets of specific biological processes such as X-chromosome inactivation and genomic imprinting (Jones, 2012).

Recently, the CpG-island centric view has been challenged, as methods to measure DNA methylation have been improved with the development of array and sequencing technology. There has been an accumulation of evidence that other genomic regions also have important roles in development and disease progression. A large portion (~80%) of methylation differences among different tissues in human and mouse has been reported to not occur in CpG islands, but 'CpG island shores', located up to 2kb distant from CpG islands (Irizarry et al., 2009). Besides the differential methylation involved in normal tissue development, most differentially methylated regions (DMRs)

that distinguish human colon cancer from normal colon tissues, human induced pluripotent stem (iPS) cells from fibroblast, and murine hematopoietic stem and progenitor cells (HSPCs) are located in CpG island shores (Doi et al., 2009; Hansen et al., 2011; Irizarry et al., 2009; Ji et al., 2010). Interestingly, the tissue-specific, cancer-specific and reprogramming-specific DMRs (tDMR, cDMR, and rDMR, respectively) showed statistically significant overlap with each other, indicating that there may be a core set of genomic regions targeted for epigenetic regulation during normal development and disease progression. In addition, DNA methylation at CpG island shores showed strong inverse correlation with gene expression, while DNA methylation at CpG islands did not show a statistically significant correlation with gene expression in these data sets, suggesting functional importance of CpG island shores in diverse biological contexts (Doi et al., 2009; Irizarry et al., 2009; Ji et al., 2010).

In cancer epigenetics, it has long been known that hypermethylation in CpG islands is a core mechanism of cancer progression, yet various types of alterations have been observed in other genomic regions. For example, in colon cancer, hypermethylation is a major type of change in CpG island shores (Irizarry et al., 2009). A large-scale change in half the genomic regions called hypomethylated 'block' has been identified in solid tumors with increased variation in expression of genes inside the blocks (Hansen et al., 2011; Timp et al., 2014; Timp and Feinberg, 2013). Stochastic variation of DNA methylation in cDMRs distinguished solid tumors from its normal counter parts, and different stages of tumors (Hansen et al., 2011; Timp et al., 2014). Thus, epigenetic dysregulation of increased epigenetic plasticity with genetic mutation would be a critical mechanism of cancer progression (Feinberg et al., 2006; Timp and Feinberg, 2013).

Recently, this hypomethylated large domain has also been observed in ageing phenotype. Sun-exposed epidermal samples in older people (Age>65) showed the hypomethylated blocks that overlapped with blocks in colon cancer and squamous cell carcinoma (Vandiver et al., 2015). The hypomethylated blocks largely overlap with higher order genomic regions such as large organized chromatin lysine-modifications (LOCKs) or nuclear lamin-associated domains (LADs), suggesting the altered large scale DNA methylation is associated with dysregulation of the large scale genomic regions in disease progression.

Besides CpG islands and shores, gene body methylation has received a lot of attention, yet its functional importance remains to be determined. DNA methylation in gene bodies has traditionally been thought to silence repetitive DNA sequences, such as retroviruses and LINE elements (Yoder et al., 1997). Recent evidence suggests that gene body methylation positively correlates with gene expression in normal tissues and cancer samples (Kulis et al., 2012; Varley et al., 2013; Yang et al., 2014). Potential mechanisms of gene body methylation in the regulation of gene expression include effects on transcription elongation or splicing regulation (Jones, 2012; Laurent et al., 2010). However, the functional role of gene body methylation in transcription regulation is not evident yet, compared to other genomic regions.

DNA demethylation

DNA methylation is a stable covalent modification on genomic sequences, while it is also reversible. Both passive and active loss of DNA methylation can occur through

diverse biological processes. Passive DNA demethylation occurs during consecutive DNA replication in the absence of functional DNMT1 activity (Kohli and Zhang, 2013). Recent advances in identification of enzymes involved in methyl group removal from cytosine have facilitated our understanding of active DNA demethylation process. Ten-eleven translocation (TET) family enzymes, TET1, TET2, and TET3 are able to oxidize methyl cytosine and generate intermediate products including 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2010; Ito et al., 2011; Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). These oxidized 5mC intermediates undergo further removal processes: passive removal by sequential DNA replication, direct removal, or DNA repair pathway-associated removal (Kohli and Zhang, 2013). Among these, the base excision repair (BER) pathway has been actively investigated. Thymine DNA glycosylase (TDG), an enzyme involved in BER is known to have an ability to remove thymine from G-T mismatches from normal DNA context (Cortazar et al., 2007). Recently, it has been reported that TDG is required for epigenetic stability in embryonic development in mice (Cortellino et al., 2011). Since TDG has an ability to remove thymine from G-T mispair, it has been hypothesized that 5mC or oxidized 5mCs may be converted to thymine or uracil by deaminase first. Several studies have suggested that AID/APOBEC enzymes, known cytosine deaminases, play a role in deamination of the 5mC in reprogramming or embryonic development (Bhutani et al., 2010; Kumar et al., 2013; Popp et al., 2010). Yet, a controversy over the role of the deaminases in DNA demethylation still exists, due to their limited enzyme activities on modified cytosines (Kohli and Zhang, 2013). In addition to the deaminase-mediated BER, other studies have shown that TDG can directly remove 5fC and 5caC (He et al., 2011;

Maiti and Drohat, 2011). DNA demethylation is implicated in multiple biological processes including pre-implantation methylation dynamics, primordial germ cell (PGS) reprograming, maintenance of stem cell pluripotency and cancer development (Kohli and Zhang, 2013).

Methods to measure DNA methylation

Array-based (CHARM and Illumina Infinium HumanMethylation450 BeadChip)

In 2008, the comprehensive high-throughput arrays for relative methylation (CHARM) method was developed to provide the first platform to interrogate DNA methylation in a genome-wide and non-CpG island biased manner (Irizarry et al., 2008). CHARM utilizes a methylation dependent restriction enzyme, McrBC, that cleaves DNA containing methylated cytosines. Sheared genomic DNA is divided into two fractions; one for undigested control and the other for McrBC digestion. The undigested control and McrBC digested samples undergo size selection. The size selected DNA is amplified and hybridized to a tiling array that excluded isolated CpGs. The log-ratio of the signal intensities from the array of the untreated and McrBC treated samples (M-value) is measured. Since methylation status of neighboring CpGs is likely to correlate with each other, the measured M-value is averaged within a given genomic region of interest. This process is called 'Genome-weighted smoothing' that improves the accuracy and specificity of measuring methylation at CpG sites (Irizarry et al., 2008). Many studies have applied CHARM to identify genome-wide differential methylation in different model systems: tissue development, cellular reprogramming, hematopoiesis, cancer, and

behavior of social insects (Doi et al., 2009; Herb et al., 2012; Irizarry et al., 2009; Ji et al., 2010; Kim et al., 2010; Kim et al., 2011).

Illumina has endeavored to develop commercially available arrays to measure genome-wide DNA methylation, and produced Infinium HumanMethylation27 BeadChip (27K) and 450 BeadChip (450K). As implied in its name, 27K covers about 27000 CpG sites, mostly enriched in CpG islands, while 450K covers about 480000 CpG sites encompassing diverse genomic regions selected from previous studies for tDMR, cDMR, rDMR, non-CpG methylated sites, and miRNA promoter regions other than CpG islands. In this section, we will focus on 450K, the method used in this thesis. The 450K array utilizes bisulfite conversion and genotyping of the C/T polymorphism method to detect methylation at a CpG site quantitatively. Methylated cytosine remains as cytosine, but unmethylated cytosine is converted to uracil after bisulfite conversion (Clark et al., 1994; Frommer et al., 1992). Note that this method does not distinguish 5mC from 5hmC, because 5hmC remains as cytosine after bisulfite conversion (Huang et al., 2010). For the 450K array, DNA sample is treated with bisulfite, then amplified by PCR. The bisulfite treated and amplified DNA is hybridized on the 450K array that returns the measurement of methylation level at CpG sites on a probe on the array (Dedeurwaerder et al., 2011).

Sequencing-based (Bisulfite pyrosequencing and whole-genome bisulfite sequencing (WGBS))

Two different levels of sequencing technology are widely used to measure DNA methylation: bisulfite pyrosequencing for a small genomic regions and WGBS for

genome-wide level. Bisulfite pyrosequencing is based on bisulfite conversion. As explained above, methylated cytosine is not affected by bisulfite treatment, while unmethylated cytosine is converted to uracil, and eventually thymidine after PCR amplification. Step-wise incorporation of deoxynucleotide triphosphates (dNTPs) during sequence extension releases pyrophosphates, which are converted to ATP by sulphurylase. Then, luciferase converts luciferin to oxyluciferin using ATP and produces light, which will be detected by pyrosequencing machine. The intensity of the released light is proportional to the amount of the nucleotides incorporated at a single base site. The C to T ratio can be quantitatively measured by the amount of dCTP or dTTP incorporation at a cytosine of a CpG site which can be inferred from the intensity of light released (Bassil et al., 2013). Bisulfite pyrosequencing has been used to detect quantitative methylation and individual CpG sites or to validate results from array-based methods (Doi et al., 2009; Herb et al., 2012; Ji et al., 2010; Kim et al., 2010).

Next-generation sequencing technology has allowed the development of WGBS, enabling researchers to investigate quantitative DNA methylation level at CpG sites genome-wide. WGBS uses bisulfite conversion as in bisulfite pyrosequencing, but the bisulfite converted DNA undergoes next-generation sequencing instead of region specific amplification. The high throughput sequencing data returns read numbers of cytosine versus thymidine, therefore yielding a quantitative measure of DNA methylation at all CpG sites in the genome (Laird, 2010; Lister et al., 2009). Several statistical methods and software packages have been developed to analyze the WGBS data: BSmooth, Bismark, and so on. Among these packages, BSmooth offers relatively accurate measurement of DNA methylation at individual CpG sites from low coverage WGBS data by using

smoothing algorithm (Hansen et al., 2012). WGBS enables researchers to discover not only DMRs, but also non-CG methylation in stem cells and large genomic DNA methylation changes in cancer and sun exposed skin in elderly people (Hansen et al., 2011; Lister et al., 2009; Vandiver et al., 2015).

2. Hematopoiesis

Overview

Hematopoiesis is one of the best studied, but complicated developmental systems. It consists of a hierarchical process initiated by hematopoietic stem cells (HSCs) that differentiate into other hematopoietic progenitor cells and eventually produce all the mature differentiated blood lineages (Chao et al., 2008; Doulatov et al., 2012). Experimental investigation of the hematopoietic system in the mouse was pioneered by Till and McCulloch who identified a small subset of cells from the mouse bone marrow which could self-renew and form myeloerythroid colonies (Becker et al., 1963; Till and Mc, 1961). These studies facilitated other investigators to develop assays such as the in vitro clonal assay combined with fluorescence-activated cell sorting (FACS), to identify and characterize hematopoietic stem and progenitor populations in mouse and human (Chao et al., 2008; Doulatov et al., 2012). The classical model of hematopoiesis demonstrates that fully differentiated blood cells constitute two major lineages: myeloid and lymphoid. Myeloid lineage cells include granulocytes, monocytes, erythrocytes, and megakaryocytes that give rise to platelets. Lymphoid lineage cells include T, B, and natural killer (NK) cells, involved in immune responses (Doulatov et al., 2012).

Human hematopoiesis

Identification of stem and progenitor cells (HSPCs) has promoted our understanding of human hematopoiesis. Multiple different cell surface markers were identified and used for isolation of HSPCs. For example, CD34 is a well known marker for the HSPCs that posses regenerative potential (Civin et al., 1984; DiGiusto et al., 1994; Krause et al., 1996). Among these CD34+ HSPCs, additional markers such as CD90, CD38, CD45RA, CD123, and CD10 enable researchers to isolate different components of the hierarchy. Several studies have demonstrated that HSC resides in Lin-CD34+CD38-CD90+ population (Chao et al., 2008). Further investigation to identify downstream progenitors of HSC has established that multipotent progenitor (MPP) cells are contained in Lin-CD34+CD38-CD90-CD45RA- (Majeti et al., 2007). After HSC gives rise to MPPs, MPPs further differentiate into progenitors for myeloid or lymphoid lineages. In myeloid lineage differentiation, the common myeloid progenitors (CMPs) develop into either GMPs or megakaryocyte/erythrocyte progenitors (MEPs). Interleukin-3 receptor alpha chain (CD123) and CD45RA distinguish CMP, GMP and MEP: CMPs reside in Lin-CD34+CD38+CD123+CD45RA-, GMPs in Lin-CD34+CD38+CD123+CD45RA+, and MEPs in Lin-CD34+CD38+CD123-CD45RA- (Chao et al., 2008). Lymphoid lineage differentiation is more complicated, since L-MPP is able to generate lymphoid lineage cells, as well as monocytes, macrophages and dendritic cells by differentiating into GMP. This L-MPP population is contained in Lin-CD34+CD38-CD90-CD45RA+ (Doulatov et al., 2010; Goardon et al., 2011). This thesis will demonstrate DNA methylation

differences among human HSPCs and compare epigenetic plasticity between human and mouse hematopoietic development.

DNA methylation in normal hematopoiesis

Two studies have investigated genome-wide DNA methylation in mouse hematopoietic development using array or sequencing based methods (Bock et al., 2012; Ji et al., 2010). Ji et al. performed CHARM examining 4.6 million CpG sites throughout the genome for MPPs, common lymphoid progenitors (CLPs), CMPs, GMPs, and thymocyte progenitors (DN1, DN2, DN3). Global methylation changes were involved in fate decision at the myeloid or lymphoid commitment stage: decreased methylation for myeloid and gain of methylation for lymphoid commitment. This first comprehensive methylome map of hematopoietic progenitor cells in murine hematopoiesis identified potential novel regulators such as Arl4c and Jdp2, as well as previously known transcription factor for hematopoietic differentiation, Meis1. This study demonstrated DNA methylation is a core mechanism for hematopoietic development, and epigenetic plasticity regulates lineage commitment (Ji et al., 2010). In another study, Bock et al. investigated genome-wide DNA methylation of HSCs, MPP1s, MPP2s, CMPs, CLPs, GMPs, MEPs, CD4-T cells, CD8-T cells, B cells, erythrocytes, granulocytes, monocytes using reduced representation bisulfite sequencing (RRBS). They observed similar pattern of epigenetic plasticity in cell fate decision: hypermethylation in CLPs and hypomethylation in CMPs. DNA methylation has played a crucial role for silencing of genes involved in myeloid differentiation in lymphoid lineage cells and vice versa. For

example, promoters of key transcription factor (TF) for myeloid differentiation such as Tal1, or binding sites of significant myeloid TFs such as Gata1 were highly methylated in CLPs compared to CMPs. This study has demonstrated that the information from the combination of DNA methylation and gene expression data accurately inferred cellular identity of different blood cells, underscoring the value of DNA methylation in hematopoietic development (Bock et al., 2012).

For human hematopoiesis, a recent study provided genome-wide DNA methylation profile for HSPCs. This study used a nano HpaII-tiny-fragment-enrichment-by-ligation-mediated-PCR (nanoHELP) assay to investigate DNA methylation of long-term HSCs (LT-HSCs), short-term HSCs (ST-HSCs), CMPs, and MEPs. Loss of methylation has been observed when ST-HSC differentiated into CMPs, and methylation changes has been correlated with gene expression at this transition, while other commitments such as CMP to MEP transition did not show statistically significant correlation between DNA methylation and gene expression. These HSC commitment-associated methylation patterns were able to predict overall patient survival in three independent AML patient cohorts, indicating the importance of epigenetic regulation for normal hematopoietic development (Bartholdy et al., 2014). This thesis will compare our results of comprehensive methylome map of human hematopoiesis to the mouse studies and the human study.

3. AML and LSC

AML

AML is a genetically heterogeneous cancer of myeloid lineage blood cells, and characterized by an accumulation of immature myeloid lineage blood cells in bone marrow. Since AML is caused by diverse pathogenic mechanisms, it is very important to identify subgroups based on the different features of the disease such as morphology and genetic alterations (Lowenberg et al., 1999). The French-American-British (FAB) classification is the most common method to differentiate the heterogeneous disease based on the morphology of leukemic blasts, indicating the degree of differentiation (Table 1.1) (Bennett et al., 1976, 1985). Specific cytogenetic abnormalities such as translocations and chromosome rearrangements correlate with particular FAB subtypes (Table 1.1). These cytogenetic lesions have been used in prognosis to predict clinical outcomes and relapse rate (Table 1.2) (Byrd et al., 2002; Grimwade et al., 1998; Slovak et al., 2000). In addition to cytogenetic abnormalities, other genetic mutations play an important role in leukemogenesis, as about a half of AML cases do not harbor a cytogenetic lesion (Dohner, 2007; Lowenberg et al., 1999). The Cancer Genome Atlas (TCGA) has provided a public resource of genomic map of over 200 AML patients by performing either whole-genome sequencing or whole-exome sequencing. This comprehensive sequencing for a large cohort of AML patients revealed the mutational landscape of AML. Interestingly, AML genomes have relatively fewer genetic mutations compared to other solid tumors, 13 mutations on average, with only 5 in genes recurrently mutated in AML. Genetic alterations were classified to nine different categories based on their biological functions of genes harboring the alterations: TF fusions (18% of cases), NPM1 mutation (27%), tumor suppressors (16%), DNA methylation enzymes (44%), activated signaling genes (59%), myeloid TFs (22%),

chromatin modifiers (30%), cohesion-complex genes (13%) and spliceosome-complex genes (14%). TF-fusion includes PML-RARA, MYH11-CBFB, RUNX1-RUNX1T1, and PICALM-MLLT10; TP53, WT1, and PHF6 for tumor suppressors; DNMT3A, DNMT3B, DNMT1, TET1, TET2, IDH1, and IDH2 for DNA methylation enzymes; FLT3, KIT, KRAS/NRAS, PTPs, and other Tyr or Ser-Thr kinases; RUNX1, CEBPA, and other myeloid TFs for myeloid TFs; MLL-X fusions, MLL-PTD, NUP98-NSD1, ASXL1, EZH2, KDM6A, and other modifiers for chromatin modifying enzymes. This study has suggested common mutations such as DNMT3A, NPM1, CEBPA, IDH1/2 and RUNX1, which were mutually exclusive to TF-fusion might be involved in the initiation of AML (Cancer Genome Atlas Research, 2013).

LSC model

LSC model postulates that AML is organized as a hierarchy like normal hematopoiesis, in which LSCs give rise to leukemic blast cells like HSCs give rise to normal progenitors and differentiated cells. In the 1990s, Dick's group identified that a small subset of CD34+CD38- cells were uniquely able to transplant AML into immune deficient mice (Bonnet and Dick, 1997; Lapidot et al., 1994). These observations lead to the hypothesis that LSCs possess increased self-renewal capacity, which enables LSCs to maintain and propagate the disease by generating bulk cancer cells (Kreso and Dick, 2014). Later, improved xenotransplantation models have revealed that LSC activity can be identified in other subpopulations as well such as CD34+CD38+ or CD34- (Eppert et al., 2011; Goardon et al., 2011; Martelli et al., 2010; Sarry et al., 2011; Taussig et al.,

2008; Taussig et al., 2010). Recently, surface makers other than CD34 and CD38 were identified to enrich for LSCs among heterogeneous cells. C-type lectin-like molecule-1 (CLL-1) was expressed in one third of CD34+CD38- compartment of 29 AML patients (van Rhenen et al., 2007). CD96 was highly expressed in CD34+CD38- compartment of about two thirds of AML patients (Hosen et al., 2007). T-cell Ig mucin-3 (TIM3), was elevated in LSC fraction compared to normal HSCs (Jan et al., 2011; Kikushige et al., 2010). CD47 was highly expressed in LSC, and expression of it protected LSCs from being phagocytosed by macrophages (Majeti et al., 2009b). CD25 and CD32 were also identified as novel marker for LSCs (Saito et al., 2010). Even though, many studies have reported a variety of surface markers for LSC, heterogeneous expression of these markers in patients suggests a complex immunophenotype that cannot be applied universally in AML.

Recently, several studies have investigated genome-wide gene expression profiles of LSCs compared to HSCs or leukemia progenitor cells (LPCs) (de Jonge et al., 2011; Eppert et al., 2011; Gentles et al., 2010; Majeti et al., 2009a). Majeti et al. performed the first genome-wide gene expression analysis of LSCs compared with HSCs. They identified 3005 differentially expressed genes, enriched in pathways such as Wnt signaling, MAP kinase signaling, adherence junction, ribosome, and T cell receptor signaling (Majeti et al., 2009a). Gentles et al. identified 52 genes which distinguished CD34+CD38- LSCs from CD34+CD38+ LPCs through genome-wide gene expression analysis. This study showed 52 genes which were associated with overall, event-free, and relapse-free survival and with therapeutic response (Gentles et al., 2010). De Jonge et al. compared CD34+ fractions with the CD34- subfraction of AML patients and CD34+

normal progenitor compartment, and found that the top 50 CD34+ specific genes were able to predict overall survival of AML patients (de Jonge et al., 2011). These studies used the cell surface markers CD34 and CD38 to define LSC compartment, following the traditional LSC model. Eppert et al. have defined LSCs functionally using a xenograft assay. They sorted AML cells into 4 different fractions based on CD34 and CD38 expression first, then performed a xenograft assay on these fractions from 16 AML patients. They compared the gene expression profile of the functionally validated LSC to non-LSC or HSCs. The analysis demonstrated that LSC and HSC share a core transcriptional program, indicating the commonality between the two populations would be derived from 'stemness' property. The genes related to the stemness were associated with clinical outcome (Eppert et al., 2011).

Cell of origin

A number of both mouse and human studies have investigated the cell of origin in AML. Mouse studies have typically utilized retroviral oncogene transduction or knock-in models to explore this question and have generally led to the conclusion that committed progenitors, in particular CMP and/or GMP, serve as the cell of origin for most AML models. In one study of MN1-induced AML, retroviral transduction of single CMP, but not GMP or HSC, resulted in the development of AML, indicating tight restriction of transformation by this oncogene (Heuser et al., 2011). In a second study using a mouse model of MLL-AF9 AML, the cell of origin influenced biological properties such as gene expression, epigenetics, and drug responses (Krivtsov et al., 2013). Both of these studies

highlight the significance of this question for leukemogenesis and potential therapies. In contrast to mouse models, inferring the cell of origin in human leukemia is only possible based on features of the disease. Studies investigating the cell of origin of human AML using surface immunophenotype and gene expression originally suggested AML LSC arise from HSC (Kreso and Dick, 2014). However, a more recent study that compared genome-wide gene expression and surface markers of LSCs to those of normal HSPCs suggests that LSCs arise from more committed progenitors, including L-MPP and GMP (Goardon et al., 2011). In blast-crisis chronic myeloid leukemia (CML), GMPs with activation of beta-catenin from patients showed increased self-renewal and leukemogenic potential (Jamieson et al., 2004). Notably, three studies have recently reported that leukemogenic mutations existed in HSC, called 'pre-leukemic HSC' that underwent further clonal evolution to give rise to AML LSC. These studies have demonstrated a hierarchy among genetic mutations during clonal evolution of the pre-leukemic HSCs. For example, mutations in epigenome modifying enzymes such as DNMT3A, IDH1/2 have occurred earlier than mutations of genes involved in activated signaling such as FLT3 and NPM1 (Corces-Zimmerman et al., 2014; Jan et al., 2012; Shlush et al., 2014). These two studies suggest that the pre-leukemic HSCs which harbor the early occurring genetic mutations is a cell of origin in AML, which may lead to disease relapse after remission.

4. DNA methylation in AML

Somatic mutations in epigenome modifying enzymes in AML

Dysregulation of the epigenome is a common feature in AML, as indicated by the recent discoveries that a number of epigenome modifying genes are mutated in AML. These genes include several involved in the regulation of DNA methylation such as IDH1/2, DNMT3A, and TET2, and modulation of chromatin modifications such as ASXL1, EZH2, and others (Abdel-Wahab et al., 2012; Cancer Genome Atlas Research, 2013; Ley et al., 2010; Yan et al., 2011). Beyond somatic mutations in these epigenome modifying factors, characterization of DNA methylation in bulk AML cells has revealed great heterogeneity among patient cases. Figueroa et al. examined ~350 AML patient samples using HpaII tiny fragment enrichment by ligation-mediated PCR (HELP) interrogating ~14000 unique gene loci. This study identified 16 distinct clusters among the patients based on DNA methylation profile, which some of the clusters were associated with particular genetic aberrations such as mutations in CEBPA, NPM1, AML1-ETO, CBFb-MYH11, and PML-RARA (Figueroa et al., 2010b). This study showed the first epigenetically distinct subtypes in AML, associated with genetic alterations. However, this study was done before genetic mutations in epigenetic modifiers were identified. The mutations in epigenome modifying enzymes induce aberrant methylation in the AML cells. In particular, AML with IDH1 or IDH2 mutations was associated with globally increased DNA methylation (Cancer Genome Atlas Research, 2013; Figueroa et al., 2010a). MLL fusion or mutations in NPM1, DNMT3A, or FLT3 were associated with decreased DNA methylation. It is interesting to observe that about a half of AML patients (44%) harbored a genetic mutation in DNA methylation enzymes, suggesting a critical role of this process in leukemogenesis (Cancer Genome Atlas Research, 2013). Recent studies have investigated how genetic mutations

in DNA methylation enzymes play an important role in leukemic transformation. Challen et al. has shown loss of function of DNMT3A disturbed HSC differentiation over serial transplantation, while the LT-HSC compartment expanded (Challen et al., 2012). DNMT3A mutations have been associated with overexpression of HSC specific genes, such as HOXA and HOXB genes in AML patients (Yan et al., 2011). As mentioned previously, TET enzymes are known to be involved in DNA demethylation by producing 5hmC intermediates. It has been reported that bone marrow samples from patients with myeloid malignancies with TET2 mutations showed decreased 5hmC with hypomethylation compared to normal controls, and the interruption of TET2 function in mouse model has displayed myeloid-skewed differentiation of HSCs. (Ko et al., 2010; Moran-Crusio et al., 2011; Quivoron et al., 2011). TET2 mutation has been implicated in clonal hematopoiesis in elderly individuals (Busque et al., 2012; Genovese et al., 2014; Jaiswal et al., 2014; Xie et al., 2014). Several studies identified recurrent mutations in IDH1 and IDH2 in AML patients and their function in leukemogenesis (Figueroa et al., 2010a; Marcucci et al., 2010; Mardis et al., 2009; Ward et al., 2010). IDH1/IDH2 mutations displayed a neomorphic enzymatic activity generating 2-hydroxyglutarate (2-HG) from α-ketoglutarate (Dang et al., 2009; Ward et al., 2010). 2-HG, oncometabolite, has been shown to inhibit TET enzyme activity, and induce promoter hypermethylation (Figueroa et al., 2010a; Shih et al., 2012). All the studies have demonstrated the significant role of the DNA methylation enzymes in regulation of epigenome in hematopoietic differentiation and disease development.

Genetic mutations in chromatin modifying enzymes play a significant role in leukemogenesis. For example, loss of function mutations of ASXL1 causes a genome-

wide loss of H3K27me3 and collaborates with oncogenes to promote leukemogenesis (Abdel-Wahab et al., 2012). The role of EZH2 mutation in myeloid malignancies is complicated, as both loss of function and gain of function mutations in EZH2, a histone lysine methyltransferase and a member of a PRC2 complex have been implicated in leukemogenesis (Ernst et al., 2010; Lund et al., 2014)

Clinical implication of epigenetics in AML patients

Mutations in DNMT3A, IDH1, IDH2, and TET2 have been linked to clinical prognosis such as risk stratification and therapeutic responses in AML patients (Shih et al., 2012). DNMT3A mutation has been associated with adverse overall survival in intermediate-risk group patients (Ley et al., 2010; Patel et al., 2012). TET2 mutation has shown adverse overall survival in AML patients with intermediate-risk, while IDH1 or IDH2 mutation, which frequently co-occurred with NPM1 mutations, has shown favorable clinical outcome (Patel et al., 2012).

In addition to the association of genetic mutation in DNA methylation enzymes with clinical outcome, DNA methylation itself has been indicated as a prognostic marker. Figueroa et al. showed an association of distinct methylation clusters of AML patients to overall survival and demonstrated that 15 genes with aberrant DNA methylation could predict overall survival of AML patients (Figueroa et al., 2010b). Furthermore, quantitative DNA methylation has successfully predicted clinical outcome in AML patients (Bullinger et al., 2010). Deneberg et al. has reported that hypermethylation at polycomb group (PcG) target genes was associated with favorable clinical outcome in cytogenetically normal AML (CN-AML) (Deneberg et al., 2011).

Besides the prognostic power of DNA methylation, the alteration of DNA methylome of myeloid malignancies has been a target of therapy, as it is reversible. Azacitidine and decitabine, DNMT inhibitors, are widely used to improve clinical outcome in AML (Estey, 2013). Recent studies have reported that drugs regulating chromatin modification could be an effective therapy for AML patients. For example, suppression of bromodomain-containing 4 (BRD4), which recognizes acetylated lysine on histone, by the small molecule inhibitor, JQ1, demonstrated anti-leukemic effects (Valent and Zuber, 2014; Zuber et al., 2011). Small molecule inhibitors of DOT1L, a telomeric silencing 1-like histone 3 lysine 79 (H3K79) methyltransferase, have shown therapeutic effects against MLL-fusion AML (Daigle et al., 2013; Daigle et al., 2011). The epigenetic therapy is an active area of research and clinical trials for AML patients.

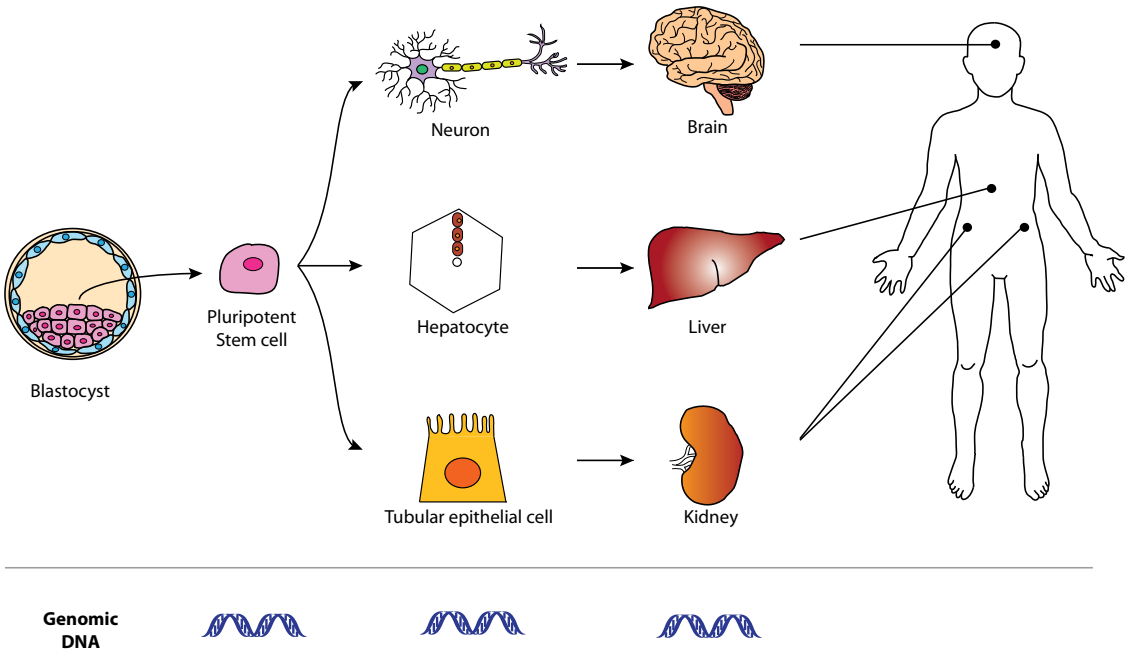


Figure 1.1. The role of epigenetics in multicellular organism.

Table 1.1. French-American-British (FAB) classification.

Type	Name	Cytogenetics
M0	Undifferentiated acute myeloblastic leukemia	
M1	Acute myeloblastic leukemia with minimal maturation	
M2	Acute myeloblastic leukemia with maturation	t(8;21)(q22;q22), t(6;9)
M3	Acute promyelocytic leukemia (APL)	t(15;17)
M4	Acute myelomonocytic leukemia	inv(16)(p13q22), del(16q)
M4eo	Acute myelomonocytic leukemia with eosinophilia	inv(16), t(16;16)
M5	Acute monocytic leukemia	del (11q), t(9;11), t(11;19)
M6	Acute erythroid leuekmia	
M7	Acute megakaryoblastic leukemia	t(1;22)

Table 1.2. Cytogenetics and prognosis.

Risk groups	Cytogenetics	5-year survival	Relapse rate
Good	t(8;21), t(15;17), inv(16)	70%	33%
Intermediate	Normal, +8, +21, +22, del(7q), Abnormal 11q23	48%	50%
Poor	-5, -7, del(5q), Abnormal 3q, Complex cytogenetics	15%	78%

Chapter2

Epigenetic signature of leukemia stem cell

Summary

AML is a hematopoietic malignancy, composed of a hierarchy that LSCs give rise into Blast cells. Since LSC is clonally related to Blasts, we hypothesized that particular epigenetic features would endow the distinct capacity of LSC, which can initiate and propagate the disease. Here, we first demonstrated epigenetic differences between LSCs and Blasts by performing genome-wide methylation analysis. We identified 84 DMRs in 70 genes, so called LSC epigenetic signature, which have shown differential methylation and expression. We found that *HOXA* cluster genes were enriched in LSC epigenetic signature, suggesting a critical role of these genes to confer the unique ability of LSC. The LSC epigenetic signature was partially dependent on genetic mutation in upstream regulators and epigenome modifying enzymes, yet about a half of it was independent of genetic mutation. The LSC epigenetic signature showed prognostic power in both DNA methylation and gene expression data sets, independent of previously known clinical factors such as age. These results provide the first evidence for epigenetic variation between LSC and their blast progeny in AML, and moreover, demonstrate that DMRs define prognostic subgroups of AML.

Results

AML LSC and Blasts Exhibit Epigenetic Differences That Define an LSC Epigenetic Signature

To formally investigate epigenetic differences between LSC and blast progeny, we sought to identify DMRs between functionally-defined AML LSC-enriched populations and their downstream non-engrafting blasts from a cohort of 15 primary patient samples. We obtained samples from 15 AML patients (Tables 2.1 and 2.2) and isolated subpopulations based on the expression of CD34 and CD38 including: Lin-CD34+CD38-, Lin-CD34+CD38+, and Lin-CD34- (Figure 2.1). We then performed comprehensive genome-scale DNA methylation analysis using the Illumina Infinium HumanMethylation450 bead chip array. While AML LSC were originally described to be exclusively contained in the CD34+CD38- subpopulation, recent reports have indicated that leukemia-initiating cells can also be detected in multiple compartments including both the CD34+CD38+ and CD34- subpopulations, although usually at lower frequencies (Eppert et al., 2011; Goardon et al., 2011; Sarry et al., 2011).

In order to identify LSC and blast populations, we conducted xenotransplantation assays on all three CD34/CD38 subpopulations from each of the 15 AML cases (Table 2.3). Similar to other reports, leukemic engraftment was observed from at least one subpopulation in 10 out of 15 AML patients. As expected, LSC activity dramatically decreased following the immunophenotypic hierarchy with 64.3% of CD34+CD38-, 46.7% of CD34+CD38+, and 26.7% of CD34- subpopulations engrafting *in vivo* (Table 2.3). To identify epigenetic markers of functional LSC, we performed DMR analysis between the 20 LSC-containing (engrafting) and 24 blast-containing (non-engrafting) fractions (hereafter termed “LSC” and “Blast”). The analysis identified 3030 DMRs, of

which 91.4% were hypomethylated in LSC (Table 2.4, see Appendix1 and Table 2.5). These DMRs were further classified according to their global genomic location including: islands (regions with a GC content greater than 50% and an observed/expected CpG ratio of more than 0.6), shores (regions within 2kb of an island), shelves (regions 2 to 4kb away from an island), and open sea (isolated CpG sites in the genome without a specific designation). These DMRs were nearly evenly distributed in CpG islands (27.8%) and open seas (29%), (Table 2.5). In addition, the DMRs strongly correlated with gene expression at CpG islands and open seas, whereas most hypomethylated DMRs in the engrafting populations were associated with transcriptional up-regulation of associated genes (Figure 2.2).

We next sought to integrate DNA methylation with gene expression analysis to identify an LSC epigenetic signature by extracting genes which passed a DMR *p* value <0.01 cutoff and exhibited >0.5 $\log_2$ fold gene expression change between the LSC and Blast populations, with an inverse relationship between gene expression and DNA methylation within 2kb of the transcriptional start site (TSS). We excluded gene body DMRs, as there was no statistically significant positive correlation in AML or normal hematopoiesis comparisons (Figure 2.3). We applied a minimum absolute value $\log_2$ 0.5 fold gene expression cutoff, similarly to our previous LSC gene expression signature using the same microarray platform (Gentles et al., 2010). With these parameters, we identified 84 regions of 70 unique genes exhibiting differential methylation and gene expression in LSC compared to Blasts (Table 2.6).

We compared our LSC epigenetic signature to the LSC gene expression signatures from previous studies (Eppert et al., 2011; Gentles et al., 2010). Only six out

of 70 genes were found in these earlier studies, suggesting most of the genes identified here comprise a novel signature for LSC defined first by DMR analysis and refined by gene expression differences. One gene in this signature, *REC8*, which encodes a kleisin family protein that is associated with the cohesin complex, was hypomethylated and transcriptionally up-regulated in LSC (Figure 2.4a and Table 2.6). Notably, mutations of components of the cohesin complex have been identified in AML and other tumor types (Losada, 2014; Thol et al., 2014). We speculate that hypomethylation and increased expression of *REC8* in LSC might be related to cohesin complex activity in LSC. We also identified *HOXA5, HOXA6, HOXA7, HOXA9,* and *HOXA10* in the LSC epigenetic signature (Figure 2.4b-d and Table 2.6). These *HOXA* cluster genes were hypomethylated and highly expressed in LSC (Table 2.6). Notably, *HOXA9* showed hypomethylation and increased expression in LSC (Figure 2.4b and Table 2.6), and aberrant expression of *HOXA9* is known to be involved in increased proliferation of HSPCs and leukemogenesis, suggesting a critical role in LSC activity (Chung et al., 2006; Lehnertz et al., 2014; Takeda et al., 2006; Thorsteinsdottir et al., 2002).

Because the MLL subtype is itself associated with changes in expression of members of the *HOXA* gene cluster (Drabkin et al., 2002; Milne et al., 2002), we performed a second DMR analysis excluding the 5 LSC populations from the 2 MLL patients in our cohort. We observed substantial overlap between the sets of DMRs without MLL cases and with all samples. For the key LSC epigenetic signature, 81 of 84 DMRs, including the *HOXA* genes, were present after removal of the MLL cases (Table 2.7). Considering all DMRs with $p < 0.01$ (not just the LSC signature), there was 77%

overlap (Table 2.7). Thus, the presence of the MLL subtype was not a confounding variable in defining the LSC epigenetic signature.

The LSC Epigenetic Signature is Partially Dependent on Underlying Somatic Mutations

In order to identify important pathways and upstream regulators of LSC activity, we utilized Ingenuity Pathway Analysis (IPA). The most significantly enriched pathway was fatty acid α oxidation (Table 2.8), and inhibitors of this pathway have been previously shown to induce apoptosis of leukemia cells (Samudio et al., 2010). Ingenuity upstream regulator analysis identified *NPM1*, *ASXL1*, and *KAT6A* as the most significant upstream regulators of the LSC epigenetic signature genes, primarily through regulation of *HOXA* genes including *HOXA5, HOXA6, HOXA7*, *HOXA9,* and *HOXA10* (Table 2.9). Significantly, all three of these upstream regulators have been found to be mutated in AML and likely serve as driver genes (Cancer Genome Atlas Research, 2013). In particular, mutations in *ASXL1* and *NPM1* have been shown to cooperate with HOX genes to initiate leukemia by enhancing self-renewal and proliferation of hematopoietic progenitors (Abdel-Wahab et al., 2012; Vassiliou et al., 2011). Consistent with this, we observed that *NPM1* mutation was associated with decreased methylation and increased expression of *HOXA5, HOXA6, HOXA7, HOXA9,* and *HOXA10* compared to *NPM1* wild-type samples in the TCGA cohort (Figure 2.5).

We then sought to investigate the LSC epigenetic signature for its association with AML mutations in the TCGA cohort (Figure 2.6). The TCGA cohort consists of 200 AML patient samples with associated DNA methylation, gene expression, and full genotyping from genome/exome sequencing (Cancer Genome Atlas Research, 2013).

First, we identified the epigenetic signatures associated with individual AML mutations by performing DMR analysis between wild-type and mutant patient samples (Figure 2.6). The mutations tested included epigenome modifying enzymes such as *DNMT3A*, *IDH1/2,* and *TET1/2*, and upstream regulators of our LSC epigenetic signature, *NPM1* and *ASXL1* (Figure 2.6). *KAT6A* was not included as there was no patient who had this mutation among the patients investigated on methylation arrays. Next, we examined the overlap between the mutation-associated DMRs and our LSC epigenetic signature (Figure 2.6). Each LSC epigenetic signature gene was classified into three categories: (1) *upstream regulator-associated* if differentially methylated in association with any mutation in upstream regulators; (2) *epigenome modifying enzyme-associated* if differentially methylated in association with any mutation in epigenetic enzymes; or (3) *mutation-independent* if it was not differentially methylated in association with either upstream regulator or epigenome modifying enzyme (Figure 2.6 and Table 2.10). Of the 84 LSC DMRs, 28 (33.3%) and 27 (32.1%) were associated with upstream regulator or epigenome modifying enzyme mutations, respectively (Figure 2.6 and Table 2.10). However, 40 DMRs (47.6%) including *HOXA7* and *HOXA9* were mutation-independent targets (Figure 2.6 and Table 2.10). It should be noted that some of the LSC differentially methylated genes, including *HOXA7* and *HOXA9*, have multiple DMRs regulated by different mechanisms. For example, *HOXA7* has 4 DMRs in the LSC epigenetic signature; one associated with mutation in *NPM1*, two associated with mutation in *DNMT3A*, *TET1,* and *NPM1*, and one mutation-independent (Table 2.10). Therefore, we annotated each DMR in those genes differently with DMR numbering such as *HOXA9/DMR1* (Figure 2.6 and Table 2.10). A small subset (11 signatures) of upstream

regulator and epigenetic enzyme associated LSC epigenetic signatures overlapped, including *REC8, HOXA6,* and *HOXA7* (Figure 2.6 and Table 2.10). This analysis showed that all the *HOXA* genes are epigenetically regulated by at least one upstream regulator, and *HOXA6, HOXA7/DMR2,* and *HOXA7/DMR3* are common targets of both upstream regulators and epigenetic enzymes (Figure 2.6 and Table 2.10), and all of these changes involved DNA hypomethylation. In addition, hypomethylation of *HOXA7/DMR1* occurred independently of mutations (Figure 2.6 and Table 2.10). Together, these results suggest that overexpression of *HOXA* genes mediated by DNA hypomethylation is a core mechanism for LSC activity.

The LSC Epigenetic Signature is Associated with Overall Survival in Human AML

We hypothesized that if the LSC epigenetic signature reflected key drivers of the functional differences between LSC and Blasts, then this signature should be associated with clinical outcomes in human AML. First, we tested the association between the LSC epigenetic signature and overall survival in the DNA methylation data from the TCGA AML cohort (Cancer Genome Atlas Research, 2013). To assign each TCGA patient to an LSC-like or Blast-like category, we calculated scores of each TCGA sample based on the probability of being closer to either LSC or Blasts. A comparable number of samples were assigned to each category by this method (99 for Blast-like and 93 for LSC-like). In univariate survival analysis, the LSC-like group showed worse outcome compared to the Blast-like group (hazard ratio (HR) =2.3, (95% confidence interval (CI) =1.6-3.4); p=1.07 x10^{-5}) (Figure 2.7a). The LSC-like vs Blast-like stratification remained associated with overall survival in multivariate analysis together with other known prognostic factors such as age (considered as a continuous variable), cytogenetic risk (assessed as

high vs low risk and intermediate vs low risk), *NPM1,* and *FLT3* mutations (HR=1.9, (95% CI= 1.2-2.9); *p*=0.003; Table 2.11).

Next, we tested the association between expression of LSC epigenetic signature genes and clinical outcome using four different cohorts including TCGA (Cancer Genome Atlas Research, 2013), a cohort of normal karyotype patients (Dufour et al., 2010; Metzeler et al., 2008), and two cohorts of mixed karyotype patients (Valk et al., 2004; Wilson et al., 2006; Wouters et al., 2009). These cohorts consist of a total of 776 AML patients treated on different clinical protocols that also exhibited distinct biological characteristics(Gentles et al., 2010). We observed a strong correlation between the relative expression of LSC epigenetic signature genes and overall survival in the TCGA cohort (correlation=0.49; p=4 $x10^{-13}$; Figure 2.8). The more highly expressed a gene was in LSC compared to Blasts, the more robust its association with worse overall survival. In all four cohorts, the overall expression level of the signature genes was significantly associated with overall survival, with higher expression associated with worse clinical (Table 2.12). This association remained significant in multivariate Cox regression including age (continuous), cytogenetic risk, *NPM1,* and *FLT3* mutations (HR=1.7, (95% CI, 1.0-2.7); *p*= 0.03; Table 2.11). Similar results were observed for the three other cohorts in univariate and multivariate analyses (Figures 2.7c-e, Tables 2.12 and 2.13).

Finally, we tested if mutations in epigenetic enzymes such as *DNMT3A*, *IDH1/2*, *TET2,* and *ASXL1* affected the prognostic impact of the LSC epigenetic signature in the TCGA cohort. As described previously, mutation in *DNMT3A*, but none of the other genes, was associated with patient overall survival (Table 2.14). Multivariate survival analysis including *DNMT3A* mutation showed that our LSC epigenetic signature

remained independently associated with clinical outcome in both the DNA methylation and gene expression data from TCGA, even when incorporating cytogenetic risk group (Table 2.15), as well as within the intermediate cytogenetic risk group alone (Table 2.16). Overall, these results demonstrate that the LSC epigenetic signature defined by DNA methylation and gene expression is associated with overall survival in human AML.

Discussion

The cancer stem cell (CSC) model was originally proposed based on observations from human AML in which only subpopulations of leukemia-initiating or LSC possessed engraftment potential (Bonnet and Dick, 1997; Lapidot et al., 1994). According to this model, the LSC give rise to downstream Blasts that lack critical stem cell properties. As LSC and their non-engrafting Blast progeny are clonally related, a major implication of this leukemia stem cell model is that their functional properties must be due to epigenetic differences. Here, we provide such evidence by characterizing global DNA methylation features of LSC defined by xenotransplantation of AML subpopulations, compared to non-engrafting Blast cells, demonstrating that AML LSC exhibit global hypomethylation compared to non-LSC Blast cells. Integrating DNA methylation and gene expression analysis, we identified 84 regions of 70 genes as the LSC epigenetic signature. 64 of these 70 genes were not reported in previous gene expression studies for LSC (the exceptions being *CD34, SH3BP5, RBPMS, LTB, MS4A3,* and *VNN1*) (Eppert et al., 2011; Gentles et al., 2010). Most of the LSC epigenetic signature was *mutation-independent*, not associated with mutations in upstream regulators or epigenome-modifying enzymes suggesting that leukemogenesis may converge on these primary epigenetic signatures We

also identified some *mutation-associated* epigenetically dysregulated genes, including *REC8* and *HOXA7*. Together, these epigenetic signatures represent potential therapeutic targets regardless of the different types of the underlying mutations present in individual AML cases. Furthermore, the LSC epigenetic signature was prognostic of patient overall survival independently of known survival predictors such as age and cytogenetic abnormalities, emphasizing its functional importance.

Apart from its prognostic effect, the LSC epigenetic signature represents a molecular target that may improve patient survival and prevent relapse. Recently, epigenetic therapy with hypomethylating agents azacytidine and decitabine has been approved for the treatment of AML. Randomized trials demonstrated improved overall survival compared to chemotherapy, but also indicated limited effect on relapse rate in high-risk AML (Estey, 2013). Our results indicate that LSC are relatively hypomethylated compared to Blasts, suggesting that they may be less effectively targeted by hypomethylating agents, possibly accounting for their limited efficacy in relapse-free survival. It would be of great interest to see how the LSC epigenetic signature is affected by these drugs.

More specifically, this LSC epigenetic signature was markedly enriched for members of the *HOXA* cluster, suggesting this cluster is a key driver of LSC function. The *HOXA* cluster has been implicated as a key regulator of hematopoiesis and myeloid malignancy (Alharbi et al., 2013). In particular, *HOXA9* is known to be involved in increased proliferation of HSPC and leukemogenesis (Thorsteinsdottir et al., 2002), even occurring as a fusion oncogene in rare cases (Nakamura et al., 1996). Moreover, increased expression of *HOXA9* has been found to be an adverse prognostic factor in

AML (Golub et al., 1999). Other *HOXA* family members including posterior (*HOXA7, HOXA9,* and *HOXA10)* and anterior (*HOXA6)* members have been implicated in leukemogenesis, as overexpression of these genes in normal mouse HSPC leads to increased self-renewal, transformation, and development of myeloid malignancies (Bach et al., 2010). The functional LSC epigenetic signature provided here demonstrates that the *HOXA* family is a key driver of AML LSC that may function in imparting aberrant self-renewal.

Materials and Methods

Human Samples

Human acute myeloid leukemia (AML) samples were collected from patient peripheral blood (PB) or bone marrow (BM) at Stanford hospital, according to an IRB-approved protocol (22264), and informed consent was obtained from all subjects. PBMC or BMMC were separated with Ficoll-Paque Plus (Amersham Biosciences, Piscataway, NJ, Catalog number: 17-1440-03), and cryopreserved in 1 x freezing medium (90%FBS + 10%DMSO). All the AML experiments were conducted with cryopreserved PBMC or BMMC samples that were thawed and washed in IMDM medium containing 10% FBS.

Flow Cytometry Analysis and Cell Sorting

A battery of antibodies (Abs) was used for staining, analysis and sorting of progenitor cells from AML patient PBMCs/BMMCs, as well as lineage analysis human chimerism/engraftment (Table 2.17). Cells were either analyzed or sorted using a FACS Aria II cytometer (BD Biosciences, Franklin Lakes, NJ). Analysis of flow cytometry raw data was done with FlowJo Software (Treestar, Ashland, OR).

Xenotransplantation Assay

NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ mice (NSG) were obtained from The Jackson Laboratory (Bar Harbor, ME) and bred in a specific pathogen-free environment per Stanford Administrative Panel on Laboratory Animal Care Guidelines (Protocol 22264). Six to eight week-old adult mice were exposed to 200 rads of gamma irradiation at least two hours (up to 24 hours) prior to transplantation. Up to 500 thousand fresh-sorted AML cell subpopulation were resuspended in 30 µl of Hank's Balanced Salt Solution (HBSS) (Gibco Life Technologies, Grand Island, NY) containing 2% FBS, and injected intravenously via the tail vein using a 29-gauge needle. For each cell subpopulation, at least three technical replicates were performed by transplantation of three aliquot of cells into three mice. Around 150 mice in total were used. Neither randomization nor blinding was used for this study.

After eight weeks, mice were euthanized with CO_2 according to IRB approved protocol (22264). BM were isolated using scissors and needle flashing, then underwent hypotonic red cell lysis using ACK (Ammonium-Chloride-Potassium) lysing buffer (Life Technologies, Grand Island, NY, Catalog# A10492). BMMCs were stained with Ab combinations (Table 2.17) on ice for 30 minutes, and dead cells were excluded by propidium iodide (PI) staining. Human myeloid engraftment (hCD45+CD33+) and lymphoid engraftment (hCD45+CD19+) were analyzed on flow as described before.

Illumina Infinium Human Methylation 450 Bead Array Assay

Genomic DNA from each sample was purified using the MasterPure DNA purification kit (Epicentre) according to the manufacturer's protocol. The genomic DNA (250-500ng) was treated with sodium bisulfate using the Zymo EZ DNA Methylation Kit (ZYMO

Research) as recommended by the manufacturer, with the alternative incubation conditions for the Illumina Infinium Methylation Assay. Converted DNA was eluted in 11ul of elution buffer. DNA methylation level was measured using Illumina Infinium HD Methylation Assay (Illumina) according to the manufacturer's specifications. Methylation array data are deposited at the Gene Expression Omnibus (GEO) with accession number GSE63409.

Illumina Infinium Human Methylation 450 Bead Array Analysis

Raw intensity files were obtained using minfi package (Aryee et al., 2014) to calculate methylation ratios (Beta values). The data was normalized using Illumina preprocessing method implemented in minfi. Several quality control measures were applied to remove arrays with low quality. Control probes were examined on the 450k array to assess several measures including bisulfite conversion, extension, hybridization, specificity and others. Next, median methylated and unmethylated signals were calculated for each arrays; no array was identified for signal values lower than 10.5. For multidimensional scaling analysis, probes containing an annotated SNP (dbSNP137) at the single-base extension or CpG sites were removed (17398 probes removed). Minfi 1.8.9 was used.

Bump hunting method previously described was applied to identify DMRs in 450k array (Aryee et al., 2014; Jaffe et al., 2012). Beta value of 0.1 (10% of methylation difference) was used as cutoff when finding DMRs. Statistical significance was assigned by permutations testing and the P-value cutoff used for downstream analysis was <0.01 that corresponded to Benjamini-Hochberg adjusted p-value <0.1 (data not shown) unless different cutoff was designated in result part. Bumphunter 1.2.0 was used. Same method

was applied to identify DMRs for the second DMR analysis of LSC vs Blast that we removed 5 LSC cases from 2 MLL patients (SU042 and SU046).

Affymetrix Microarray Expression Analysis

Total RNA was extracted from each FACS-sorted cell population using RNeasy® Plus Mini (QIAGEN, Valencia, CA, Catalog#: 74134) according to the manufacture's protocol. All RNA samples were quantified with 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), subjected to reverse transcription, two consecutive rounds of linear amplification, and production and fragmentation of biotinylated cRNA. 15μg of cRNA from each sample was hybridized to HG U133 Plus 2.0 microarrays. Hybridization and scanning were performed according to the manufacture's instruction (Affymetrix). This step was performed at the PAN center of Stanford University. Data were normalized by GC robust multi-array average method and analyzed on R/Bioconductor. SU042 CD34+38+ was removed from further analysis due to low quality. SU001 was excluded, as the samples from this patient were not included on expression array (GEO GSE63270).

Sanger Sequencing to Detect AML Mutations

Genomic DNA was extracted from patient BMMC or PBMC using QIAmp DNA Mini Kit (QIAGEN, Valencia, CA, Catalog#: 51304) according to the manufacture's instruction. PCR primers were designed to cover exon 3-11 of TET2, exon 4 of IDH1/2, and exon 7-23 of DNMT3A (Table 2.18). The PCR reaction premix consists of 1x of OneTaq 2x Master Mix (NEB, Ipswich, MA, Catalog#: M0482L), 0.2μM forward and reverse primers respectively, and 10ng (up to 100ng) genomic DNA as template. The reaction was under the condition of 95ºC initial denaturation for 30 seconds, 45 cycles of

extension containing 94ºC for 30 seconds, 56ºC for 1 minute (or as necessary) and 72ºC for 1 minute, and a final extension at 72ºC 5 minutes. The PCR products were concentrated with PCR purification kit (QIAGEN, Valencia, CA, Catalog#: 28106), then submitted to Sequentech (Mountain View, CA) for sequencing of both forward and reverse directions using 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA) according to the manufacturer's instruction. The sequencing data was analyzed using Sequencher 5.1 (Gene Codes Corporation, Ann Arbor, MI), and single-nucleotide polymorphism (SNP) was excluded by checking NCBI website before getting the final mutation results.

Survival Analysis

Survival analysis was performed to assess the association of LSC DNA methylation and gene expression signatures with clinical outcome (overall survival) in 4 different cohorts. For DNA methylation data set (TCGA), patients were separated into two groups; LSC-like and Blast-like based on the methylation profile of each individual. Survival was compared between the two groups using the *coxph* function in R (survival package 2.37), with significance assessed by log-rank test. For gene expression, the genes in the LSC epigenetic signature were identified in expression datasets for which survival outcomes were available. The first principal component of their expression levels was computed, and patients were stratified as "high" or "low" relative to its median value. Survival differences between the groups were assessed by log-rank test. In multivariate analyses, age was incorporated as a continuous variable, mutations were coded as present/absent (1/0), and assessment of cytogenetic risk was treated as individual groups and done for

intermediate vs low-risk and High vs low risk (Figure 2.9). Analysis was also performed within intermediate risk groups.

Statistical Analysis

To assign cell identity of LSC/Blast to TCGA samples, mean methylation value of each LSC epigenetic signature (84 DMRs) for LSC/Blast (methylation profile) was retrieved and standard deviation of the mean value for each signature was calculated. Then scores (probability density values as log value) for each TCGA sample regarding LSC and Blast profile was calculated using dnorm function with the mean and standard deviation calculated in previous step. Maximum value of scores between the ones regarding LSC and Blast methylation profile was chosen, and then cell identity assigned.

Bioinformatics Analysis

QIAGEN's Ingenuity IPA (Ingenuity® Systems, www.ingenuity.com) was performed for pathway analysis.

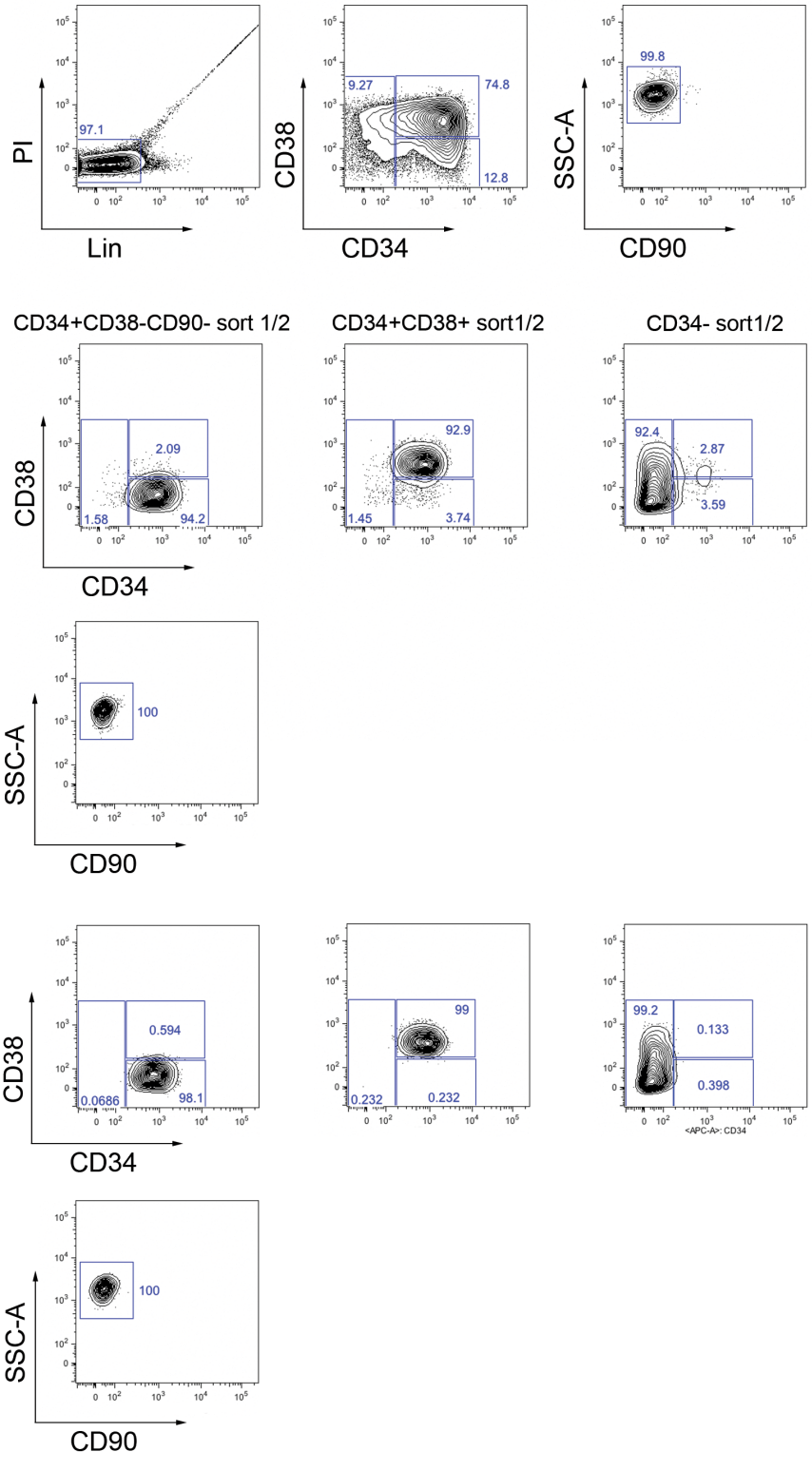

PI
Lin
97.1
CD38
CD34
9.27
74.8
12.8
SSC-A
CD90
99.8
CD34+CD38-CD90- sort 1/2
CD34+CD38+ sort1/2
CD34- sort1/2
2.09
1.58
94.2
92.9
1.45
3.74
92.4
2.87
3.59
100
0.594
0.0686
98.1
99
0.232
0.232
99.2
0.133
0.398
<APC-A>: CD34

Figure 2.1. Pre-sort and post-sort FACS analysis of subpopulations from human AML. Top panel: FACS-sorting scheme of three immunophenotypically defined subpopulations from human AML samples. Other panels: Two rounds of post-sort analysis to check the purity of sorting.

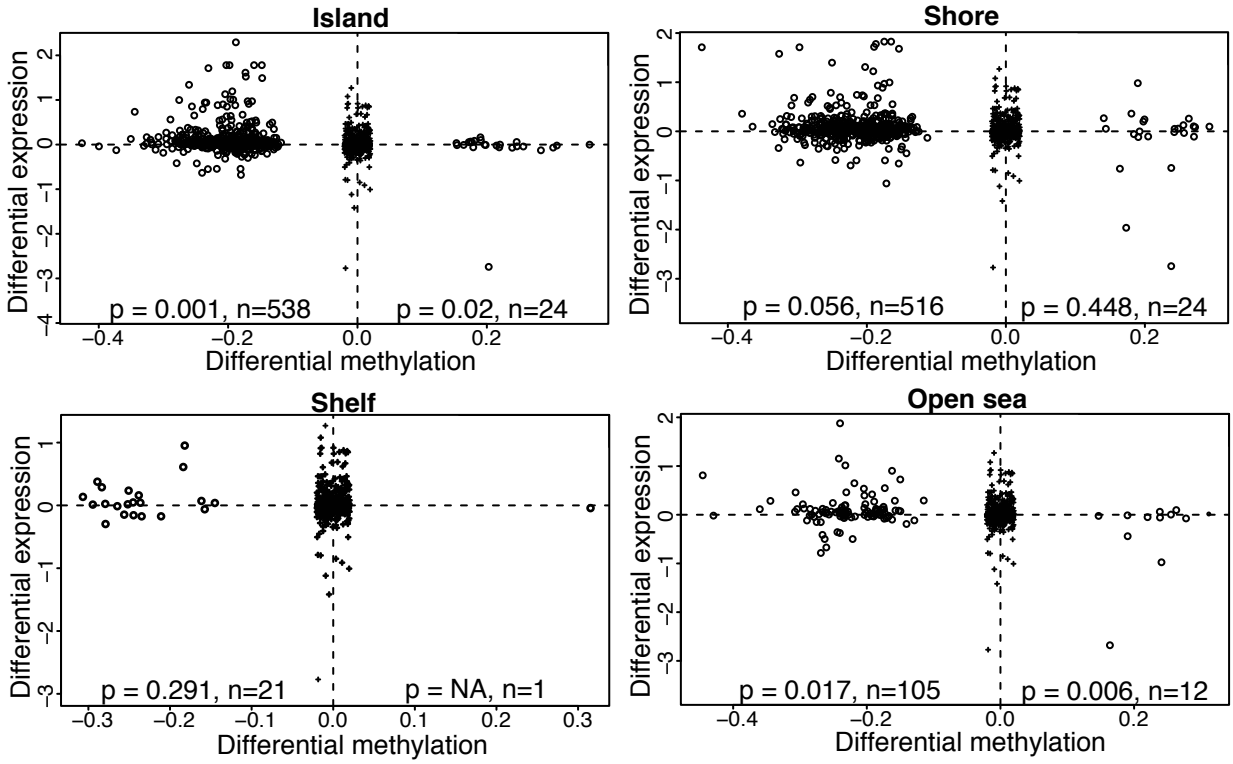
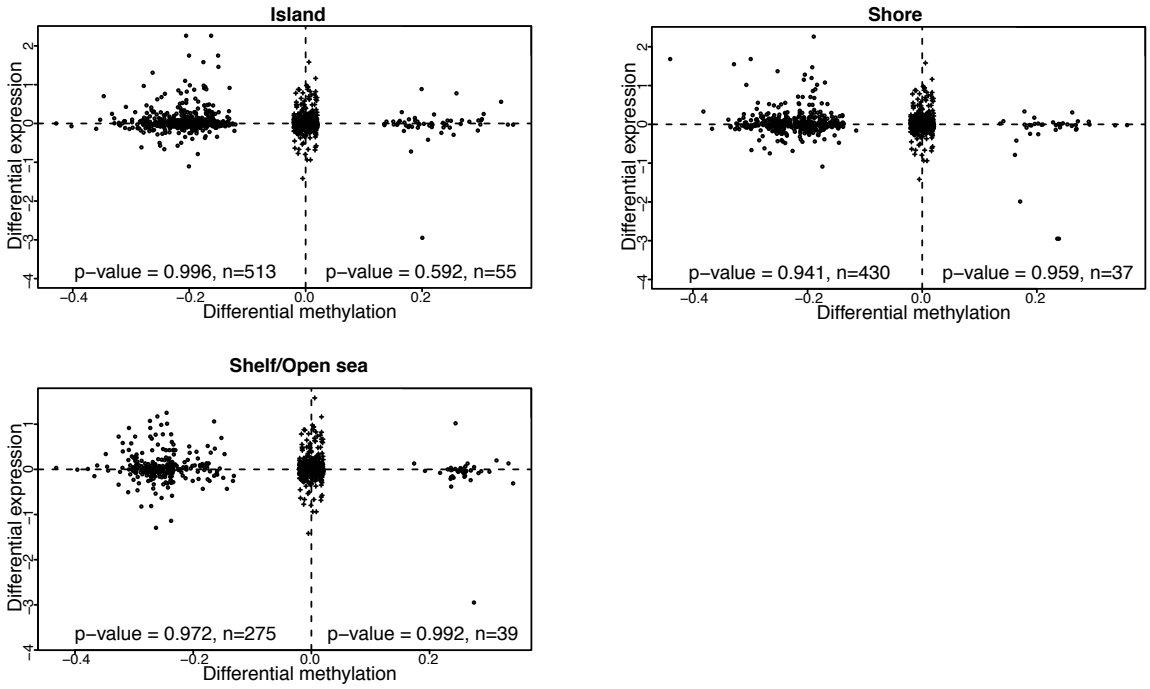

Island
Differential expression
Differential methylation
p = 0.001, n=538
p = 0.02, n=24
Shore
p = 0.056, n=516
p = 0.448, n=24
Shelf
p = 0.291, n=21
p = NA, n=1
Open sea
p = 0.017, n=105
p = 0.006, n=12

Figure 2.2. Gene expression inversely correlates with DMRs at CpG island and open sea. Engrafting (LSC) and non-engrafting (blast) subpopulations from primary AML cases were profiled for DNA methylation and gene expression to identify differentially methylated regions (DMRs) and differentially expressed genes between these two groups. DMRs that are located within 2kb of gene transcriptional start sites (TSSs - black dots) were classified into 4 groups according to their distance relative to a CpG island: island, shore, shelf, and open sea. DMRs located further than 2kb away from TSSs are denoted as black pluses. Log_2 ratios of differential expression were plotted against differential methylation (all values are blast compared to LSC). Wilcoxon rank-sum tests were performed to test the null hypothesis that the expression differences for the hypo- or hypermethylated DMRs within 2kb of gene TSSs (black dots) showed stronger inverse correlation than the expression differences of the random DMRs that are located further than 2kb of TSSs (black pluses). Random DMRs were shown in the middles of DNA methylation axis regardless of their methylation differences.

a


Island
Differential expression
p−value = 0.996, n=513
p−value = 0.592, n=55
Differential methylation
Shore
Differential expression
p−value = 0.941, n=430
p−value = 0.959, n=37
Differential methylation
Shelf/Open sea
Differential expression
p−value = 0.972, n=275
p−value = 0.992, n=39
Differential methylation

b

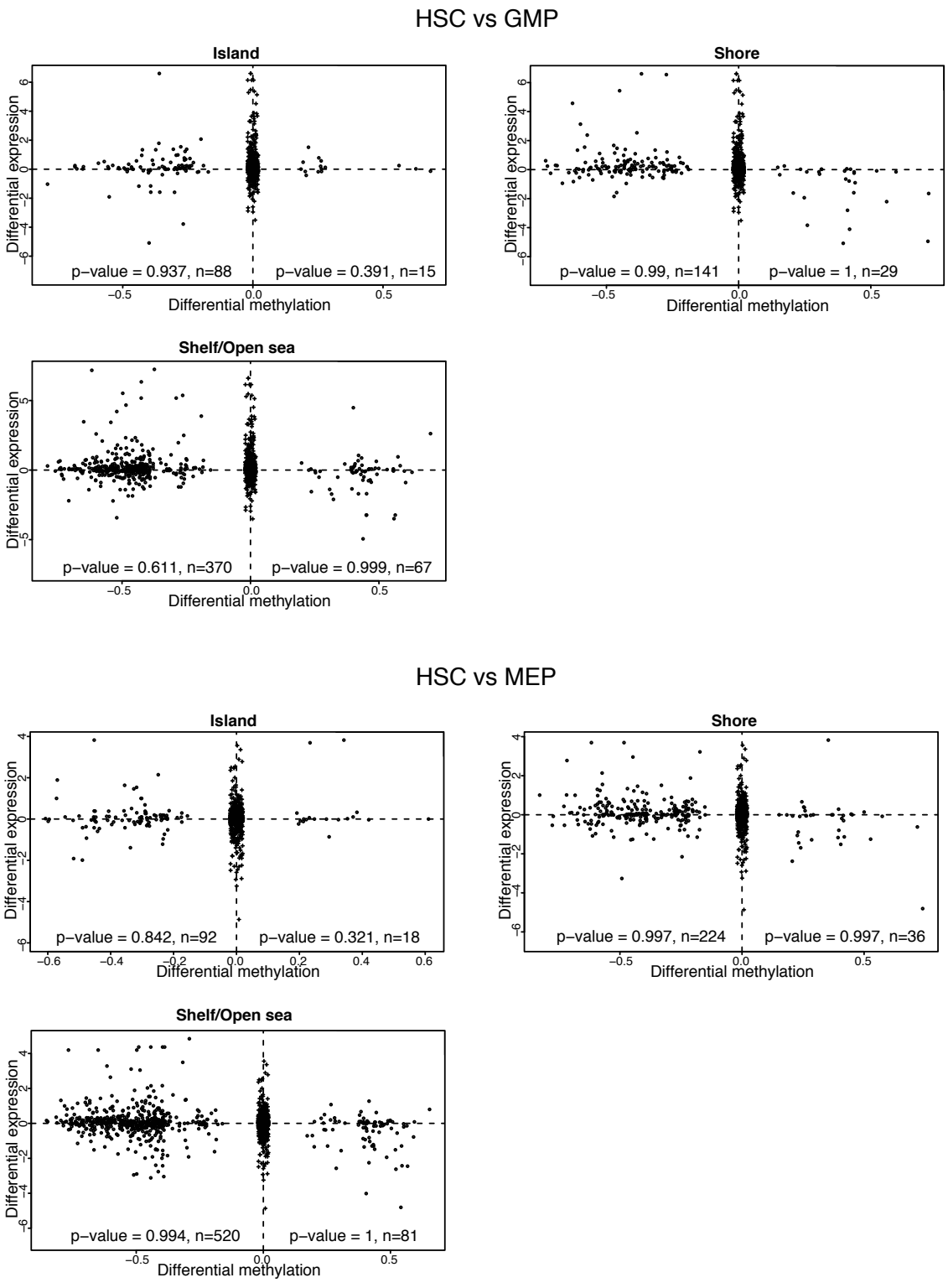

HSC vs GMP
Island
Shore
Shelf/Open sea
Differential expression
Differential methylation
p−value = 0.937, n=88
p−value = 0.391, n=15
p−value = 0.99, n=141
p−value = 1, n=29
p−value = 0.611, n=370
p−value = 0.999, n=67
HSC vs MEP
Island
Shore
Shelf/Open sea
Differential expression
Differential methylation
p−value = 0.842, n=92
p−value = 0.321, n=18
p−value = 0.997, n=224
p−value = 0.997, n=36
p−value = 0.994, n=520
p−value = 1, n=81

Figure 2.3. Gene body methylation doesn't show statistically significant positive correlation with gene expression. DMRs that are located in gene body (TSS to transcription end site (TES)) were classified into three groups according to their distance relative to a CpG island: island, shore, shelf/open sea. Random DMRs that don't locate in gene body are denoted as black pluses. Log_2 ratios of differential expression were plotted against differential methylation. Wilcoxon rank-sum tests were performed to test the null hypothesis that the expression differences for the hypo- or hypermethylated DMRs located in gene body (black dots) showed stronger positive correlation than the expression differences of the random DMRs that do not locate in gene body (black pluses). (**a**) LSC vs Blast. All values for DNA methylation and gene expression are from LSC-Blast. (**b**) Normal hematopoiesis. HSC vs GMP and HSC vs MEP are shown. All values for DNA methylation and gene expression are from group2 – group1 for group1 vs group2 comparisons.

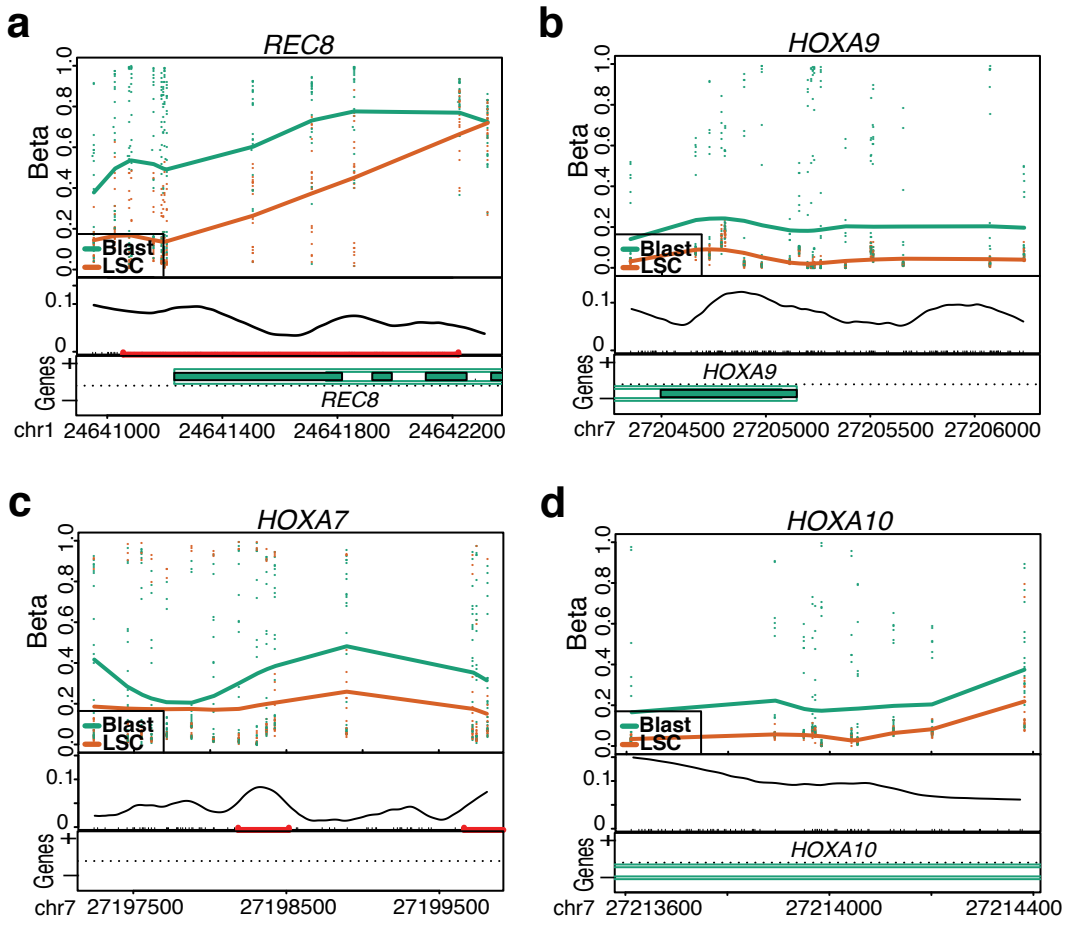
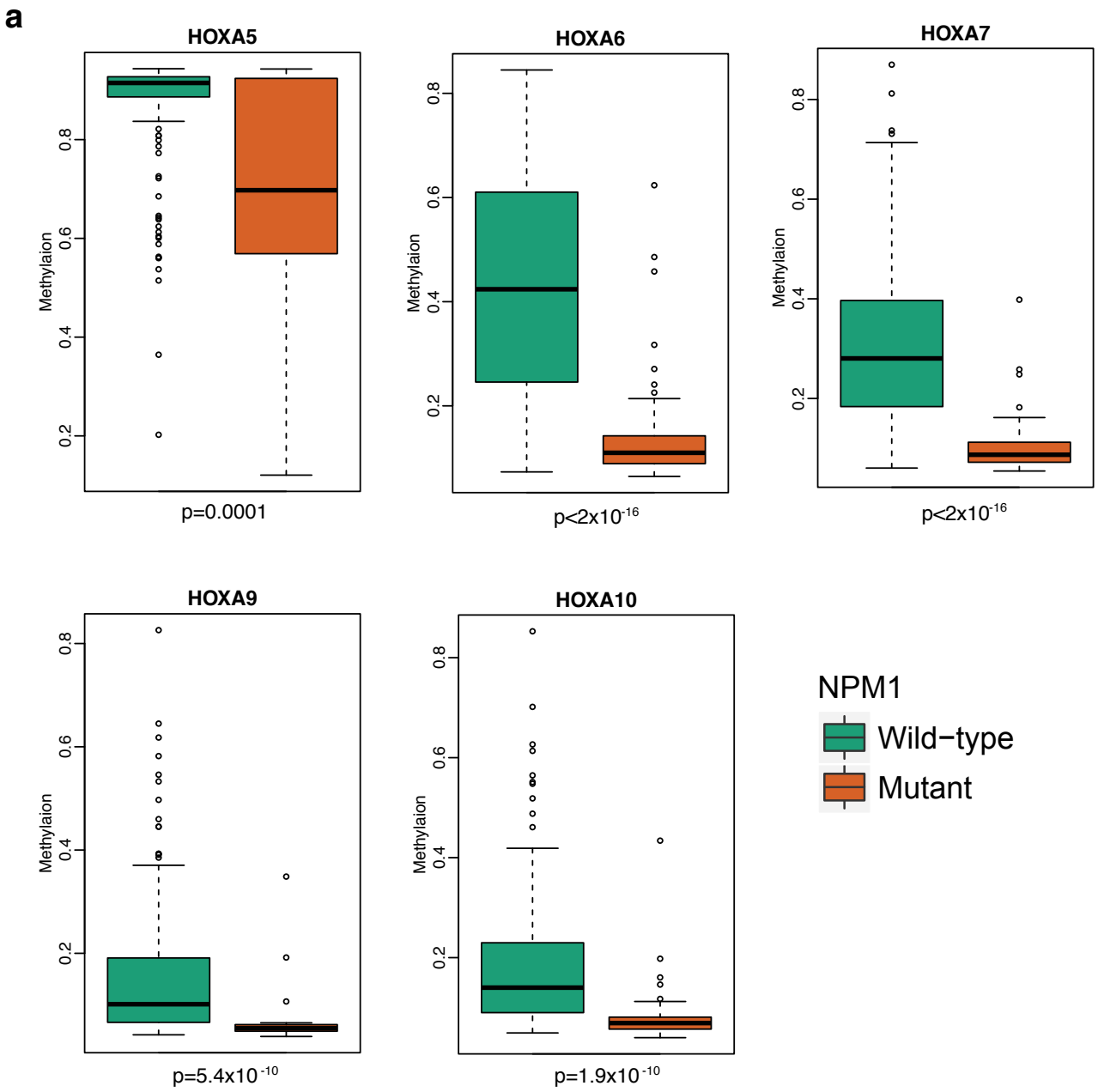


Figure 2.4. AML LSC and Blasts exhibit epigenetic differences that define an LSC epigenetic signature. Plots of differentially methylated regions (DMRs) indicating genomic loci for *REC8* (**a**), *HOXA9* (**b**), *HOXA7* (**c**) and *HOXA10* (**d**), four LSC epigenetic signature genes that are hypomethylated and upregulated in LSC. Top: level of CpG methylation (beta) of each sample for the region; Middle: CpG density (curve), CpG sites (black tick marks), CpG islands (red lines); Bottom: gene annotation.


a
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
Methylaion
p=0.0001
p<2x10-16
p<2x10-16
p=5.4x10-10
p=1.9x10-10
NPM1
Wild−type
Mutant

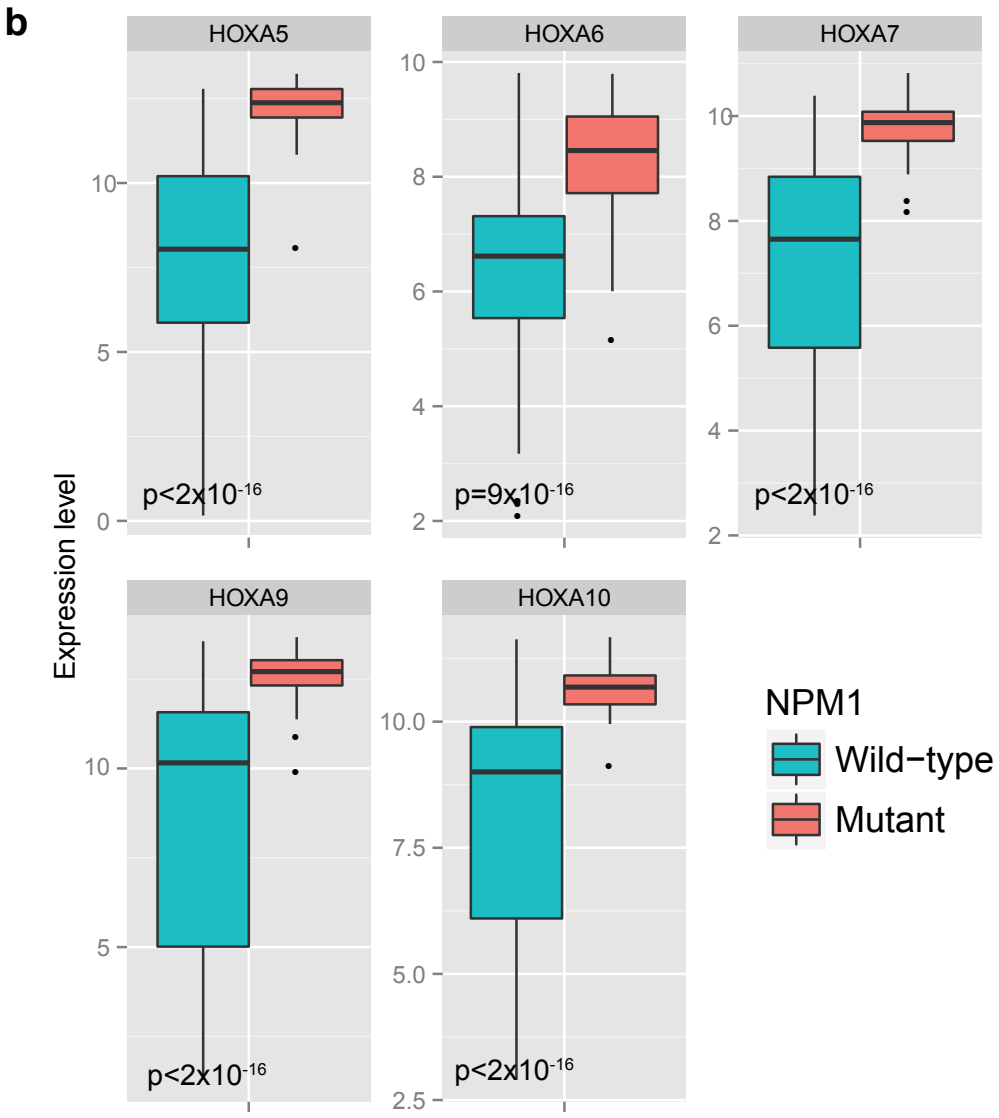

b
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
Expression level
p<2x10-16
p=9x10-16
p<2x10-16
p<2x10-16
p<2x10-16
NPM1
Wild−type
Mutant

Figure 2.5. *NPM1* mutation is associated with decreased methylation and increased expression of *HOXA* genes. (**a**) Box plots show methylation level for *NPM1* mutants and wild-type samples for DMRs for *HOXA5*, *HOXA6*, *HOXA7*, *HOXA9,* and *HOXA10* in the TCGA dataset. t-test assuming unequal variance was performed to look at statistical significance of the association between *NPM1* mutation and methylation. DNA methylation of all the *HOXA* genes was significantly associated with *NPM1* mutation. (**b**) Box plots show gene expression (Log_2 value) for *NPM1* mutants and wild-type samples for *HOXA5*, *HOXA6*, *HOXA7*, *HOXA9,* and *HOXA10* in the TCGA dataset. t-test assuming unequal variance showed *NPM1* mutation highly correlated with increased expression of all the *HOXA* genes tested.

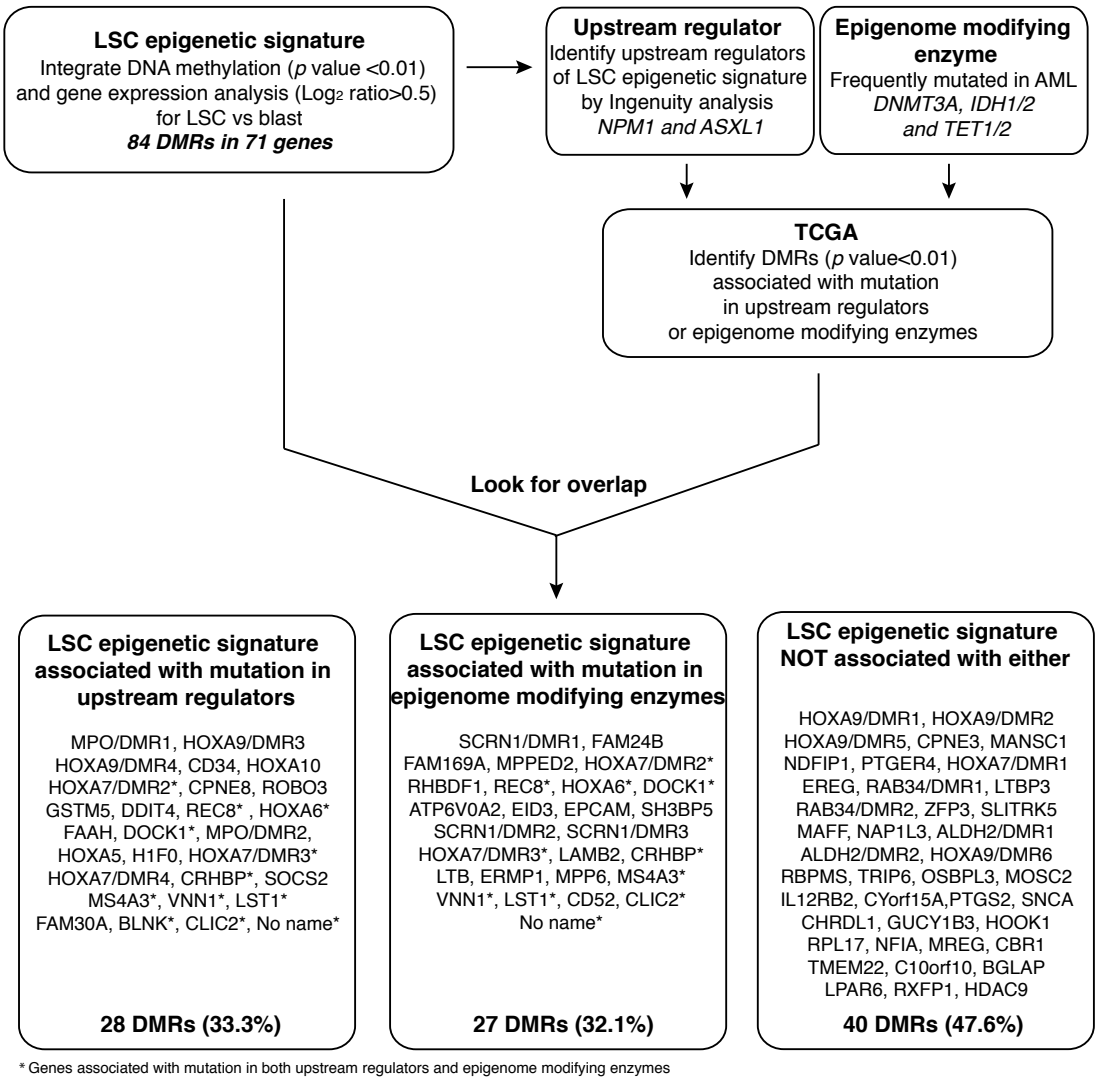


Figure 2.6. The LSC epigenetic signature is partially dependent on underlying somatic mutations. Shown is a schematic flow chart of mutation association analysis. We compared our LSC epigenetic signature to the mutation specific DMRs obtained from TCGA data set. The LSC epigenetic signature was classified into three different groups, and each DMR is shown in this diagram. Note that several genes such as *HOXA9* have multiple DMRs and different DMRs in one gene are annotated with DMR number such as *HOXA9/DMR1*.

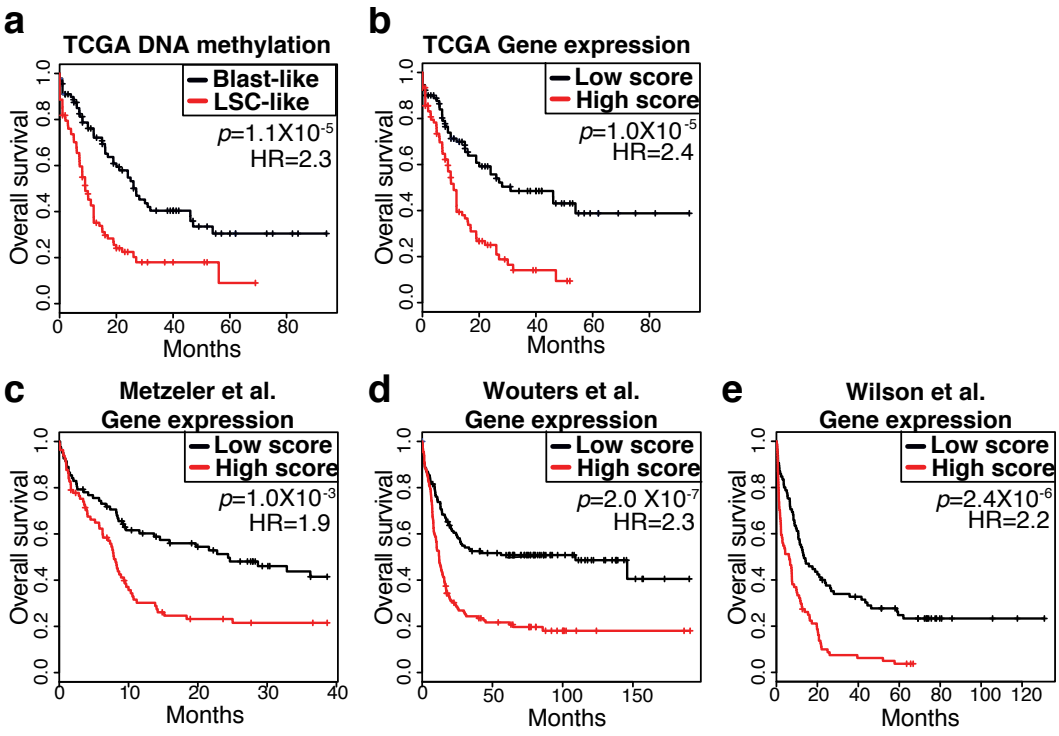

a
TCGA DNA methylation
Blast-like
LSC-like
p=1.1X10-5
HR=2.3
Overall survival
Months
b
TCGA Gene expression
Low score
High score
p=1.0X10-5
HR=2.4
Overall survival
Months
c
Metzeler et al.
Gene expression
Low score
High score
p=1.0X10-3
HR=1.9
Overall survival
Months
d
Wouters et al.
Gene expression
Low score
High score
p=2.0 X10-7
HR=2.3
Overall survival
Months
e
Wilson et al.
Gene expression
Low score
High score
p=2.4X10-6
HR=2.2
Overall survival
Months

Figure 2.7. The LSC epigenetic signature is associated with overall survival in human AML. (**a**) TCGA samples were classified as LSC-like or Blast-like based on DNA methylation alone by generating methylation profiles of the LSC and Blast populations, and then calculating scores of each sample based on the probability of being closer to either LSC or Blast. Kaplan-Meier survival analysis was then applied to these groups as indicated. Statistical significance was determined by the Log-rank test (n=192; 93 LSC-like and 99 Blast-like patients). (**b-e**) Expression of the LSC epigenetic signature genes was combined to create an LSC score, which was then calculated in AML samples from four independent cohorts including TCGA (n=182; 91 high and 91 low score patients) (**b**) Metzeler et al (n=163; 81 high and 82 low score patients) (**c**), Wouters et al (n=262; 131 high and 131 low score patients) (**d**), and Wilson et al (n=169, 84 high and 85 low score patients) (**e**). In each cohort, patients were classified into high and low groups based on the median value. Kaplan-Meier survival analysis was then applied to these groups as indicated. Statistical significance was determined by the Log-rank test.

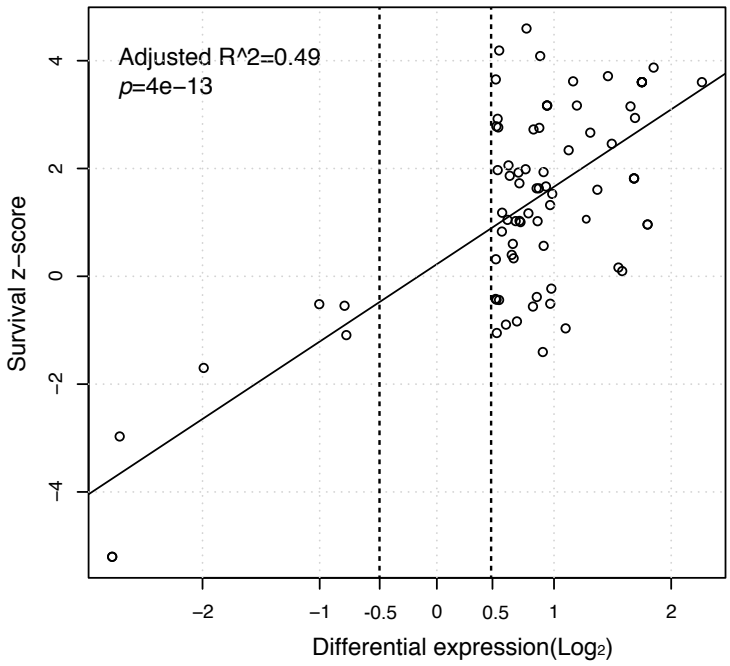


Figure 2.8. The gene expression of the LSC epigenetic signature highly correlates with clinical outcome in the TCGA dataset. Each dot represents an LSC epigenetic signature gene. Survival z-score was plotted against $\log_2$ ratio of differential expression of the LSC epigenetic signature genes in TCGA.

a

```
## DNA methylation analysis and survival
stime <- ifelse(tpd$vital_status=="DECEASED",tpd$days_to_death,tpd$days_to_last_followup)
stime[stime%in%c("[Not Available]","[Not Applicable]")]<-NA
stime<-as.numeric(stime)
event<-tpd$vital_status=="DECEASED"
age<-tpd$age_at_initial_pathologic_diagnosis
prog<-tpd$acute_myeloid_leukemia_calgb_cytogenetics_risk_category
prog[prog=="[Not Available]"]<-NA
prog<-factor(prog,levels=c("Favorable","Intermediate/Normal","Poor"),labels=c("F","I","P"))
library(survival)
summary(coxph(Surv(stime,event)~group+age+prog+Flt3+Npm1)
summary(coxph(Surv(stime,event)~group+age+prog+Flt3+Npm1+tpd$dnmt3a)
```

b

```
## GEP analysis and survival

dmrexp = read.delim("TCGA_AML_newDMR_p001_fc05.eigengenes.pcl", stringsAsFactors=FALSE)
amlinfo2 = merge(amlinfo,dmrexp, by="Array")
medexp = median(amlinfo2$DMR_p0.01_fc0.5, na.rm=TRUE)
amlinfo2$medexp = 1
amlinfo2$medexp[amlinfo2$DMR_p0.01_fc0.5>medexp] = 2
summary(coxph(Surv(OS_Time,OS_Status) ~
DMR_p0.01_fc0.5+NPM+FLT3+dnmt3a+Age+CALGB_cytorisk, data=amlinfo2))
summary(coxph(Surv(OS_Time,OS_Status) ~
DMR_p0.01_fc0.5+NPM+FLT3+Age+CALGB_cytorisk, data=amlinfo2))
```

Figure 2.9. R script for multivariate survival analysis. (**a**) Multivariate survival analysis for DNA methylation data in TCGA. The line or a variable that show how we treated cytogenetic groups is colored in red. (**b**) Multivariate survival analysis for gene expression data in TCGA. The line or a variable that show how we treated cytogenetic groups is colored in red.

Table 2.1. Clinical features of AML patients in this study

Sample ID	Age	Gender	1°/2°	D/R	Cytogenetics	% CD34+	WHO Classification	FAB
SU001	59	F	1°	R	Normal	99	AML-not otherwise specified	M2
SU006	51	F	1°	D	Failed to grow	94	AML-not otherwise specified	M1
SU008	64	M	1°	D	Normal	3	AML-not otherwise specified	M1
SU014	59	M	1°	D	Normal	18	AML-not otherwise specified	ND
SU029	65	F	1°	D	inv(9)(p11q13)	8	AML with multilineage dysplasia without antecedent MDS	M2
SU032	47	M	1°	D	Normal	68	AML-not otherwise specified	M5
SU035	46	M	1°	D	Failed to grow	98	AML-not otherwise specified	M5
SU036	71	F	1°	D	t(8;21)	47	AML with t(8;21)(q22;q22)	ND
SU042	61	F	1°	D	t(10;11)	8	AML with 11q23 (MLL)	M5b
SU046	53	F	1°	D	t(6;11)	94	AML with 11q23 (MLL)	M5
SU056	56	M	1°	D	Complex cytogenetics	99	AML with multilineage dysplasia without antecedent MDS	M0
SU266	65	M	1°	D	inv(3)	96	AML with inv(3)(q21q26)	ND
SU267	58	M	1°	D	Normal	66	AML with multilineage dysplasia without antecedent MDS	ND
SU302	59	M	1°	D	Normal	14	AML-not otherwise specified	ND
SU306	33	F	1°	D	No analyzable metaphases	<1	AML-not otherwise specified	M5a

Abbreviations: 1°, primary; 2°, secondary; D, de novo; F, female; M, male; ND, no data; R, relapsed

Table 2.2. Genetic mutations identified

Patient ID	TET2	IDH1	IDH2	DNMT3A	FLT3 ITD	FLT3 TKD	NPM1	KIT	CEBPA
SU001	wt	wt	wt	wt	wt	nd	wt	nd	nd
SU006	wt	wt	wt	wt	wt	nd	wt	nd	nd
SU008	wt	wt	wt	wt	mut	wt	wt	nd	nd
SU014	wt	R132H	wt	wt	mut	nd	mut	nd	nd
SU029	1149FS	wt	wt	R882H	mut	nd	mut	nd	nd
SU032	Y1649C	wt	wt	wt	wt	nd	wt	nd	nd
SU035	wt	wt	wt	wt	wt	nd	wt	nd	nd
SU036	wt	wt	wt	wt	nd	nd	wt	mut	nd
SU042	wt	wt	wt	S837*	wt	nd	wt	nd	nd
SU046	wt	wt	wt	wt	wt	wt	wt	nd	Nd
SU056	wt	wt	wt	wt	wt	wt	wt	nd	wt
SU266	E1010D	wt	wt	wt	wt	wt	wt	nd	wt
SU267	wt	R132C	wt	R882H	wt	wt	wt	nd	wt
SU302	wt	wt	wt	R882H	wt	wt	mut	wt	mut
SU306	wt	wt	R140Q	ΔV149	wt	mut	mut	wt	wt

Abbreviations: FS, frameshift mutation; wt, wild type; mut, mutant; nd, no data; * stop; Δ, deletion.

Note: Sanger sequencing was performed on TET2 exon 3-11, IDH1, IDH2 exon 4, and DNMT3A exon 3-11. More details are provided in Table 2.18. For all other mutations, data are derived from clinical laboratory testing.

Table 2.3. Engraftment of AML subpopulations

Patient ID	"CD34-"	"CD34+CD38+"	"CD34+CD38-"
SU001	No	No	No
SU006	No	No	Yes
SU008	No	No	No
SU014	No	No	No
SU029	Yes	Yes	Yes
SU032	No	No	No
SU035	Yes	No	Yes
SU036	No	No	No
SU042	Yes	Yes	Yes
SU046	Yes	Yes	ND
SU056	No	Yes	Yes
SU266	No	Yes	Yes
SU267	No	Yes	Yes
SU302	No	Yes	Yes
SU306	No	No	Yes
Frequency	4/15 (26.7%)	7/15 (46.7%)	9/14 (64.3%)

Note: Yes: engrafted; No: no-engraftment; ND, no data. For SU046, there is no CD34+CD38- cell fraction.

Table 2.4. DMRs of LSC vs Blast (**See Appendix1**)

Table 2.5. Summary of LSC vs Blast DMRs

Comparisons (Group1 versus Group2)	Numbers of DMRs*		Locations of DMRs relative to CpG islands (%)			
	Group1>Group2	Group1<Group2	Islands	Shores	Shelves	Open seas
Blast vs LSC	2769	261	27.8	37.8	5.4	29

* *P* value cutoff of 0.01 was used to calculate the number of DMRs (see Methods)

Table 2.6. LSC epigenetic signature

chr	start	end	diffMethyl	island	diffexp	Gene	DMR_TSS_dist
chr17	56356470	56356963	0.200454417	Island	-2.771069849	MPO	1304
chr7	27209463	27209582	-0.190449837	Island	2.262250718	HOXA9	-194
chr7	27205200	27205262	-0.205288088	Island	1.750125912	HOXA9	-51
chr7	27203430	27203546	-0.200286397	Island	1.750125912	HOXA9	1603
chr7	27206073	27206907	-0.1623543	Island	1.750125912	HOXA9	-924
chr7	27204052	27204981	-0.150359078	Island	1.750125912	HOXA9	168
chr7	30029717	30029808	-0.232917635	Island	1.682967495	SCRN1	-402
chr1	208083913	208084071	-0.175662022	Island	1.581896622	CD34	612
chr8	87526705	87527257	-0.175070398	Island	1.492929541	CPNE3	-8
chr7	27213984	27214383	-0.149812433	Island	1.460273965	HOXA10	0
chr12	12502846	12502846	-0.262873441	Island	1.309407455	MANSC1	329
chr5	141488047	141488121	-0.196034559	Island	1.163061111	NDFIP1	314
chr10	124638756	124639630	-0.206994825	Island	1.098988409	FAM24B	0
chr5	74162602	74162809	-0.198910627	Island	0.984734984	FAM169A	0
chr11	30605787	30606026	-0.194575258	Island	0.976935267	MPPED2	-226
chr5	40681444	40681444	-0.278087288	Island	0.968009601	PTGER4	-1406
chr7	27195918	27196286	-0.172243476	Island	0.940816384	HOXA7	10
chr7	27198025	27198896	-0.165502698	Island	0.940816384	HOXA7	-1729

chr16	122031	122031	-0.238501186	Island	0.9115447	RHBDF1	562
chr4	75230391	75230615	-0.236192175	Island	0.91073247	EREG	244
chr12	39299364	39299726	-0.170515768	Island	0.881992857	CPNE8	0
chr17	27045043	27045302	-0.210388337	Island	0.87035073	RAB34	-97
chr11	65325158	65325249	-0.218616458	Island	0.852913568	LTBP3	139
chr17	27044169	27044685	-0.191399132	Island	0.851673857	RAB34	0
chr11	124747075	124747263	-0.258430121	Island	0.825061807	ROBO3	-338
chr1	110254692	110255096	-0.204217813	Island	0.820222184	GSTM5	0
chr10	74034644	74034667	-0.241705206	Island	0.765158417	DDIT4	-963
chr14	24640947	24641852	-0.346838032	Island	0.704630646	REC8	0
chr17	5000803	5001047	-0.180094485	Island	0.639449923	ZFP3	-1867
chr13	88326244	88326244	-0.254835678	Island	0.622086898	SLITRK5	-1375
chr7	27188020	27188465	-0.198478999	Island	0.611495803	HOXA6	-652
chr1	46859671	46860511	-0.204828969	Island	0.603323126	FAAH	0
chr10	128593922	128594144	-0.21139227	Island	0.532062199	DOCK1	0
chr12	124247223	124247223	-0.291619587	Island	0.531144201	ATP6V0A2	-1636
chr22	38610376	38610795	-0.197239852	Island	0.519140873	MAFF	-516
chr12	104697193	104697631	-0.159774024	Island	0.511530854	EID3	893
chrX	92928508	92928610	-0.215885822	Island	0.509900024	NAP1L3	0
chr17	56357994	56358318	0.235604062	Shore	-2.771069849	MPO	0
chr2	47597118	47597331	0.170825315	Shore	-1.989838284	EPCAM	-652

chr12	112204756	112205368	-0.177217632	Shore	1.798343931	ALDH2	-6
chr12	112203801	112204506	-0.167361987	Shore	1.798343931	ALDH2	244
chr7	27205504	27205514	-0.189312475	Shore	1.750125912	HOXA9	-355
chr3	15372726	15372965	-0.191871292	Shore	1.691364717	SH3BP5	923
chr7	30028281	30028307	-0.439758955	Shore	1.682967495	SCRN1	1008
chr7	30027454	30027454	-0.299227828	Shore	1.682967495	SCRN1	1861
chr7	27184077	27184159	-0.156088057	Shore	1.652615374	HOXA5	-794
chr22	38201496	38201848	-0.328478194	Shore	1.54987965	H1F0	-242
chr8	30243930	30243930	-0.252405503	Shore	1.369854497	RBPMS	-1888
chr7	100465051	100465833	-0.20478317	Shore	1.281381937	TRIP6	-78
chr7	27193351	27194013	-0.193087116	Shore	1.195082055	HOXA7	219
chr7	25018503	25018595	-0.16970931	Shore	0.966906711	OSBPL3	1066
chr7	27196759	27197239	-0.177214748	Shore	0.940816384	HOXA7	-463
chr1	220922046	220922217	-0.186478516	Shore	0.904301462	MOSC2	-436
chr1	67772896	67773044	-0.179679693	Shore	0.859823651	IL12RB2	2
chr3	49170496	49170794	0.161770178	Shore	-0.78786347	LAMB2	0
chrY	21728575	21728575	0.23566283	Shore	-0.772593829	CYorf15A	692
chr1	186650441	186650479	-0.254143241	Shore	0.757276969	PTGS2	-885
chr5	76249502	76250527	-0.21663642	Shore	0.713935248	CRHBP	-634
chr6	31549563	31550090	-0.20860554	Shore	0.705812972	LTB	112
chr4	90757139	90757378	-0.206564689	Shore	0.683467526	SNCA	-293

chr9	5831674	5831999	-0.28057155	Shore	0.672688294	ERMP1	-697
chrX	110039536	110039604	-0.240784813	Shore	0.667757542	CHRDL1	-543
chr4	156681475	156681475	-0.306895344	Shore	0.655209216	GUCY1B3	-1234
chr1	60280088	60280106	-0.186569105	Shore	0.648435017	HOOK1	488
chr18	47016218	47016218	-0.284177026	Shore	0.616623366	LOC729046 /// RPL17	-932
chr7	24614206	24614348	-0.286787283	Shore	0.557276298	MPP6	-1183
chr1	61549542	61549982	-0.239027386	Shore	0.555128781	NFIA	-1563
chr12	93966060	93967711	-0.165376292	Shore	0.52397313	SOCS2	0
chr2	216877276	216877750	-0.197016908	Shore	0.50372197	MREG	565
chr21	37442759	37442777	-0.191372117	Shore	0.503048275	CBR1	-423
chr3	136539328	136539328	-0.246575275	Shore	0.501386314	TMEM22	-1351
chr10	45474317	45474372	-0.183899524	shelf	0.931044445	C10orf10	-60
chr1	156211409	156211434	-0.185683842	shelf	0.589711142	BGLAP	570
chr11	59823993	59824116	0.161461095	Open sea	-2.706906127	MS4A3	14
chr6	133035379	133035379	-0.242403002	Open sea	1.849875377	VNN1	-191
chr6	31556255	31556255	-0.244242862	Open sea	1.1249658	LST1	-1279
chr13	48987165	48987165	0.238629161	Open sea	-1.003641807	LPAR6	-562
chr14	106354912	106354912	-0.234479961	Open sea	0.985697199	FAM30A	1067
chr1	26644515	26645313	-0.164618365	Open sea	0.874440839	CD52	-31
chr4	159442782	159442782	-0.447897492	Open sea	0.781456346	RXFP1	117
chr10	98031125	98031337	-0.152269885	Open sea	0.696536347	BLNK	0

chr7	47611829	47611926	-0.220983979	Open sea	0.619420183	TNS3	-918
chrX	154563852	154563968	-0.175452155	Open sea	0.518822469	CLIC2	0
chr7	18535072	18535786	-0.206752311	Open sea	0.507854096	HDAC9	0

Table 2.7. Second DMR analysis to examine confounding effect of MLL cases

	LSC epigenetic signature	DMRs (p value<0.01)
All Samples	84	3030
No MLL cases	49	1398
Overlap	73.5% (36/49)	77% (1076/1398)

Table 2.8. Ingenuity pathway analysis

Ingenuity Canonical Pathways	-log(p-value)	Molecules
Fatty Acid α-oxidation	2.8E00	ALDH2,PTGS2
Melatonin Degradation III	1.81E00	MPO
Anandamide Degradation	1.81E00	FAAH
NRF2-mediated Oxidative Stress Response	1.73E00	GSTM5,MAFF,CBR1
VDR/RXR Activation	1.59E00	BGLAP,HOXA10
Eicosanoid Signaling	1.57E00	PTGS2,PTGER4
Phenylethylamine Degradation I	1.47E00	ALDH2
HGF Signaling	1.35E00	DOCK1,PTGS2
Prostanoid Biosynthesis	1.34E00	PTGS2
Parkinson's Signaling	1.31E00	SNCA
Corticotropin Releasing Hormone Signaling	1.25E00	PTGS2,GUCY1B3
Granzyme A Signaling	1.22E00	H1F0
Glutathione Redox Reactions I	1.12E00	CLIC2
Relaxin Signaling	1.12E00	RXFP1,GUCY1B3
Aryl Hydrocarbon Receptor Signaling	1.12E00	GSTM5,NFIA
Histamine Degradation	1.09E00	ALDH2
eNOS Signaling	1.08E00	LPAR6,GUCY1B3
Tryptophan Degradation X (Mammalian, via Tryptamine)	1.08E00	ALDH2

Oxidative Ethanol Degradation III	1.08E00	ALDH2
Putrescine Degradation III	1.06E00	ALDH2
Ethanol Degradation IV	1.06E00	ALDH2
Triacylglycerol Degradation	1.03E00	FAAH
Phenylalanine Degradation IV (Mammalian, via Side Chain)	1.02E00	ALDH2
Gap Junction Signaling	1.01E00	BGLAP,GUCY1B3
IL-9 Signaling	9.95E-01	SOCS2
MIF-mediated Glucocorticoid Regulation	9.83E-01	PTGS2
Role of JAK2 in Hormone-like Cytokine Signaling	9.83E-01	SOCS2
Dopamine Degradation	9.5E-01	ALDH2
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	9.39E-01	MPO,HOXA10
Inhibition of Angiogenesis by TSP1	9.39E-01	GUCY1B3
Glutathione-mediated Detoxification	9.39E-01	GSTM5
IL-8 Signaling	9.35E-01	MPO,PTGS2
Endothelin-1 Signaling	9.35E-01	PTGS2,GUCY1B3
Sertoli Cell-Sertoli Cell Junction Signaling	9.35E-01	MPP6,GUCY1B3
ILK Signaling	9.32E-01	DOCK1,PTGS2
Ethanol Degradation II	8.99E-01	ALDH2
MIF Regulation of Innate Immunity	8.9E-01	PTGS2
FcγRIIB Signaling in B Lymphocytes	8.81E-01	BLNK
Primary Immunodeficiency Signaling	8.23E-01	BLNK

Noradrenaline and Adrenaline Degradation	8.23E-01	ALDH2
Lymphotoxin β Receptor Signaling	7.93E-01	LTB
Role of IL-17A in Arthritis	7.93E-01	PTGS2
Nur77 Signaling in T Lymphocytes	7.79E-01	HDAC9
Huntington's Disease Signaling	7.74E-01	HDAC9,SNCA
Colorectal Cancer Metastasis Signaling	7.6E-01	PTGS2,PTGER4
Phototransduction Pathway	7.59E-01	GUCY1B3
Role of JAK1 and JAK3 in γc Cytokine Signaling	7.46E-01	BLNK
Phospholipase C Signaling	7.43E-01	BLNK,HDAC9
Cell Cycle: G1/S Checkpoint Regulation	7.4E-01	HDAC9
CD40 Signaling	7.34E-01	PTGS2
Antiproliferative Role of Somatostatin Receptor 2	7.22E-01	GUCY1B3
Macropinocytosis Signaling	7.16E-01	RAB34
Role of MAPK Signaling in the Pathogenesis of Influenza	7E-01	PTGS2
IL-17 Signaling	6.94E-01	PTGS2
T Helper Cell Differentiation	6.94E-01	IL12RB2
JAK/Stat Signaling	6.94E-01	SOCS2
Glucocorticoid Receptor Signaling	6.92E-01	BGLAP,PTGS2
Growth Hormone Signaling	6.89E-01	SOCS2
Small Cell Lung Cancer Signaling	6.84E-01	PTGS2
STAT3 Pathway	6.84E-01	SOCS2

TREM1 Signaling	6.73E-01	MPO
Prolactin Signaling	6.73E-01	SOCS2
Cyclins and Cell Cycle Regulation	6.63E-01	HDAC9
Serotonin Degradation	6.63E-01	ALDH2
Superpathway of Melatonin Degradation	6.58E-01	MPO
ErbB Signaling	6.26E-01	EREG
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	6.17E-01	LTB
Crosstalk between Dendritic Cells and Natural Killer Cells	6.09E-01	LTB
FAK Signaling	6.09E-01	DOCK1
Neuregulin Signaling	5.97E-01	EREG
Chronic Myeloid Leukemia Signaling	5.97E-01	HDAC9
PPAR Signaling	5.93E-01	PTGS2
p53 Signaling	5.77E-01	HDAC9
Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes	5.73E-01	DOCK1
Telomerase Signaling	5.73E-01	HDAC9
IGF-1 Signaling	5.73E-01	SOCS2
Paxillin Signaling	5.62E-01	DOCK1
Cholecystokinin/Gastrin-mediated Signaling	5.48E-01	PTGS2
Pancreatic Adenocarcinoma Signaling	5.41E-01	PTGS2
Type I Diabetes Mellitus Signaling	5.31E-01	SOCS2
Nitric Oxide Signaling in the Cardiovascular System	5.28E-01	GUCY1B3

Gαs Signaling	5.25E-01	PTGER4
Hereditary Breast Cancer Signaling	5.12E-01	HDAC9
Gα12/13 Signaling	5.09E-01	LPAR6
14-3-3-mediated Signaling	5.06E-01	SNCA
RhoA Signaling	4.91E-01	LPAR6
LXR/RXR Activation	4.8E-01	PTGS2
PI3K/AKT Signaling	4.8E-01	PTGS2
Ovarian Cancer Signaling	4.67E-01	PTGS2
PI3K Signaling in B Lymphocytes	4.67E-01	BLNK
Sperm Motility	4.61E-01	GUCY1B3
Type II Diabetes Mellitus Signaling	4.54E-01	SOCS2
Protein Kinase A Signaling	4.53E-01	PTGS2,H1F0
IL-12 Signaling and Production in Macrophages	4.51E-01	IL12RB2
Cellular Effects of Sildenafil (Viagra)	4.44E-01	GUCY1B3
Synaptic Long Term Depression	4.23E-01	GUCY1B3
CXCR4 Signaling	4.03E-01	DOCK1
Axonal Guidance Signaling	3.99E-01	DOCK1,ROBO3
Acute Phase Response Signaling	3.81E-01	SOCS2
Dopamine-DARPP32 Feedback in cAMP Signaling	3.73E-01	GUCY1B3
Role of NFAT in Regulation of the Immune Response	3.7E-01	BLNK
Dendritic Cell Maturation	3.66E-01	LTB

B Cell Receptor Signaling	3.62E-01	BLNK
Role of NFAT in Cardiac Hypertrophy	3.54E-01	HDAC9
Calcium Signaling	3.54E-01	HDAC9
Mitochondrial Dysfunction	3.5E-01	SNCA
Agranulocyte Adhesion and Diapedesis	3.47E-01	CD34
ERK/MAPK Signaling	3.45E-01	DOCK1
mTOR Signaling	3.4E-01	DDIT4
Hepatic Fibrosis / Hepatic Stellate Cell Activation	3.29E-01	LTB
Integrin Signaling	3.27E-01	DOCK1
Actin Cytoskeleton Signaling	3E-01	DOCK1
cAMP-mediated signaling	2.97E-01	PTGER4
LPS/IL-1 Mediated Inhibition of RXR Function	2.96E-01	GSTM5
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	2.94E-01	BGLAP
G-Protein Coupled Receptor Signaling	2.48E-01	PTGER4
Xenobiotic Metabolism Signaling	2.38E-01	GSTM5
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	2.1E-01	LTB

Table 2.9. Ingenuity upstream regulator analysis

Upstream Regulator	p-value of overlap	Target molecules in dataset
ASXL1	8.18E-12	HOXA10,HOXA5,HOXA6,HOXA7,HOXA9
KAT6A	4.16E-07	HOXA5,HOXA7,HOXA9
NPM1	1.09E-06	HOXA10,HOXA5,HOXA7,HOXA9
phorbol myristate acetate	1.88E-06	BGLAP,CD52,CRHBP,DDIT4,FAAH,HOXA5,HOXA7,HOXA9,IL12RB2,LTB,MPO,PTGER4,PTGS2
KMT2A	6.98E-06	HOXA10,HOXA5,HOXA7,HOXA9
raloxifene	8.44E-06	BGLAP,BLNK,DDIT4,LAMB2,MPPED2,PTGS2
nimesulide	1.38E-05	EREG,MPO,PTGS2
1-methyl-4-phenylpyridinium	1.66E-05	DDIT4,MPO,SNCA
mir-223	2.19E-05	CD34,CRHBP,GSTM5,MREG,NFIA
EPZ004777	2.31E-05	HOXA10,HOXA9
beta-estradiol	2.73E-05	BGLAP,BLNK,DDIT4,FAAH,GUCY1B3,HOXA10,HOXA9,LAMB2,LTB,MPO,MPPED2,OSBPL3,PTGS2,RBPMS,SOCS2
diethylstilbestrol	2.93E-05	BGLAP,EREG,HOXA10,HOXA9,SOCS2,VNN1
EZH2	3.78E-05	HOXA10,HOXA6,HOXA7,HOXA9,LTB,PTGS2
PHF1	5.19E-05	HOXA10,HOXA6,HOXA9
RNF20	7.67E-05	HOXA10,HOXA9
MEN1	1.18E-04	BGLAP,HOXA7,HOXA9

paricalcitol	1.28E-04	BGLAP,PTGER4,PTGS2
bexarotene	1.53E-04	C10orf10,DDIT4,MAFF,PTGS2
PSIP1	1.60E-04	HOXA7,HOXA9
arsenite	2.08E-04	EREG,MAFF,MPO,PTGS2
HBP1	2.14E-04	H1F0,PTGS2
LIF	2.18E-04	BGLAP,CD34,EREG,PTGS2,SOCS2
IL12 (complex)	2.39E-04	FAAH,IL12RB2,LTB,PTGS2,SOCS2
staurosporine	2.58E-04	BGLAP,CPNE3,MPO,PTGS2
AHR	2.64E-04	ALDH2,CBR1,GSTM5,HDAC9,LTBP3,PTGS2
trans-hydroxytamoxifen	3.27E-04	BLNK,DDIT4,LAMB2,MPPED2
meloxicam	3.42E-04	MPO,PTGS2
3-methylcholanthrene	3.52E-04	BGLAP,GSTM5,PTGS2
beta-naphthoflavone	3.73E-04	CBR1,GSTM5,SOCS2
tretinoin	3.85E-04	BGLAP,CBR1,CD34,CD52,EPCAM,HOXA5,HOXA9,MPO,MPP6,MS4A3,PTGS2,REC8
miR-196a-5p (and other miRNAs w/seed AGGUAGU)	3.89E-04	FAM169A,HDAC9,HOOK1,HOXA5,HOXA7,HOXA9
RNF2	3.96E-04	HOXA5,HOXA7,REC8
rotenone	4.19E-04	MPO,PTGS2,SNCA
lipopolysaccharide	4.48E-04	ALDH2,EREG,FAAH,GUCY1B3,HDAC9,IL12RB2,LST1,MAFF,MPO,NDFIP1,PTGER4,PTGS2,SOCS2
5'-	5.00E-04	GUCY1B3,PTGS2

methylthioadenosine		
PLA2G4A	5.00E-04	GSTM5,PTGS2
miR-221-3p (and other miRNAs w/seed GCUACAU)	5.37E-04	CLIC2,CPNE8,DDIT4,HOXA7,NDFIP1,OSBPL3,SLITRK5
RNA polymerase II	5.75E-04	BGLAP,BLNK,HOXA7,HOXA9,PTGS2
mir-196	5.89E-04	HOXA7,HOXA9
EREG	5.89E-04	EREG,PTGS2
infliximab	6.08E-04	DDIT4,IL12RB2,PTGS2
E. coli B5 lipopolysaccharide	6.79E-04	ALDH2,IL12RB2,MPO,PTGS2,SOCS2
NFATC2	6.79E-04	IL12RB2,PTGS2,SH3BP5,SOCS2
RELB	7.03E-04	LTB,MPO,PTGS2
doxycycline	7.37E-04	HOXA10,MPO,PTGS2
ESR1	7.85E-04	DDIT4,EREG,GUCY1B3,LTB,PTGS2,SOCS2
BDKRB2	7.91E-04	PTGS2,SNCA
dexmedetomidine	7.91E-04	DDIT4,PTGS2
CBFB	1.00E-03	MPO,PTGER4,SOCS2
TNF	1.02E-03	ALDH2,BGLAP,C10orf10,EREG,HDAC9,HOXA9,LTB,MAFF,MPO,PTGS2,RBPMS,SOCS2
miR-1243 (miRNAs w/seed ACUGGAU)	1.13E-03	BLNK,NAP1L3,SNCA
CSF3	1.14E-03	CD34,HOXA7,LTB,MPO
NPC2	1.15E-03	PTGER4,PTGS2

ceruletide	1.15E-03	MPO,PTGS2
forskolin	1.15E-03	BGLAP,CRHBP,DDIT4,EREG,GSTM5,HOXA5,PTGS2
dexamethasone	1.22E-03	BGLAP,C10orf10,CBR1,CRHBP,DDIT4,EREG,GSTM5,HOXA7,LAMB2,LTB,PTGS2,SOCS2
fumonisin B1	1.28E-03	LTB,PTGS2
genistein	1.30E-03	EREG,GUCY1B3,HOXA10,OSBPL3,PTGS2
EHF	1.37E-03	BLNK,EREG,VNN1
HOXA9	1.41E-03	CD34,CLIC2,HOXA9,NFIA
DOT1L	1.42E-03	HOXA10,HOXA9
CSF2	1.46E-03	LTB,MPO,PTGER4,PTGS2,REC8,SOCS2
stearic acid	1.56E-03	PTGS2,SNCA
Histone h3	1.62E-03	BGLAP,HOXA10,HOXA5,HOXA7,HOXA9
prostaglandin A1	1.72E-03	LPAR6,PTGS2
BMP15	1.72E-03	EREG,PTGS2
beta-glycerophosphoric acid	1.72E-03	BGLAP,PTGS2
CFTR	1.81E-03	BLNK,FAAH,PTGS2
SND1	1.88E-03	EREG,PTGS2
wortmannin	1.92E-03	BGLAP,PTGER4,PTGS2,SOCS2
COL18A1	2.00E-03	CD34,DDIT4,PTGS2
JAK	2.04E-03	PTGER4,SOCS2
TRIB3	2.04E-03	BGLAP,DDIT4

SP3	2.15E-03	BGLAP,EREG,IL12RB2,PTGS2
IFNG	2.20E-03	BLNK,C10orf10,FAAH,HDAC9,IL12RB2,LAMB2,LST1,LTB,PTGS2,SOCS2
CD28	2.25E-03	GUCY1B3,IL12RB2,PTGER4,PTGS2,SOCS2
cigarette smoke	2.27E-03	GSTM5,MAFF,PTGS2,SOCS2
FOLR1	2.34E-03	EPCAM,LAMB2,MPP6
AREG	2.40E-03	EREG,PTGS2
RUNX1	2.48E-03	BGLAP,CD34,MPO
Sos	2.49E-03	DOCK1,HOOK1,LTBP3,PTGS2
FOS	2.63E-03	BGLAP,DOCK1,EREG,HOOK1,LTBP3,PTGS2
dextran sulfate	2.76E-03	GSTM5,MPO,PTGS2,REC8
3-methyladenine	2.78E-03	PTGS2,SNCA
grape seed extract	2.78E-03	MPO,PTGS2
butylated hydroxyanisol	2.78E-03	CBR1,GSTM5
thromboxane A2	2.80E-03	PTGS2
Gαq	2.80E-03	PTGS2
Jnkk	2.80E-03	PTGS2
olesoxime	2.80E-03	MPO
Pad2	2.80E-03	PTGS2
LPAR5	2.80E-03	PTGS2
PAP1	2.80E-03	PTGS2

FLT4	2.80E-03	BGLAP
Tpl2 kinase inhibitor	2.80E-03	PTGS2
Mucin	2.80E-03	PTGS2
PF-4523655	2.80E-03	DDIT4
cucurbitacin E	2.80E-03	PTGS2
dixanthogen	2.80E-03	GSTM5
EDN2	2.80E-03	PTGS2
CLN3	2.80E-03	HOOK1
SIGLEC7	2.80E-03	PTGS2
SIGLEC9	2.80E-03	PTGS2
WDR61	2.80E-03	HOXA9
ACSL4	2.80E-03	PTGS2
TFF1	2.80E-03	PTGS2
TLE6	2.80E-03	BGLAP
Pla2g2a	2.80E-03	PTGS2
LY311727	2.80E-03	PTGS2
BN 50730	2.80E-03	PTGS2
1-(1-glycero)dodeca-1,3,5,7,9-pentaene	2.80E-03	PTGS2
CGP77675	2.80E-03	PTGS2
Ro 31-7549	2.80E-03	PTGS2

arachidic acid	2.80E-03	SNCA
n-6 docosapentaenoic acid	2.80E-03	PTGS2
long-chain alcohol	2.80E-03	PTGS2
IL10	2.91E-03	C10orf10,FAAH,HDAC9,IL12RB2,PTGS2
HDAC4	2.95E-03	CHRDL1,HDAC9,PTGS2
PRKG1	2.98E-03	GUCY1B3,MPO
COMMD3-BMI1	2.98E-03	HOXA7,HOXA9
HOXA7	2.98E-03	CD34,HOXA7
NR3C2	3.03E-03	DDIT4,NFIA,PTGS2
TSC22D3	3.19E-03	BGLAP,PTGS2
miR-3976 (miRNAs w/seed AUAGAGA)	3.19E-03	CPNE3,HDAC9
miR-3186-5p (miRNAs w/seed AGGCGUC)	3.19E-03	CLIC2,FAM169A
UPF2	3.19E-03	MPPED2,PTGS2
taurine	3.19E-03	MPO,PTGS2
tetrachlorodibenzodioxin	3.21E-03	BLNK,CBR1,HDAC9,PTGS2,SOCS2
STAT6	3.22E-03	ALDH2,LTB,PTGS2,SOCS2
IFNA2	3.32E-03	IL12RB2,LPAR6,REC8,SOCS2
ERK	3.38E-03	BGLAP,EREG,MAFF,PTGS2

benzo(a)pyrene	3.38E-03	EREG,MAFF,MPO,PTGS2
Immunoglobulin	3.49E-03	LST1,MAFF,MPO,PTGS2
Histone h4	3.65E-03	BGLAP,HOXA9,PTGS2
melatonin	3.74E-03	BGLAP,MPO,PTGS2
FGF1	3.93E-03	GSTM5,LTB,PTGS2
miR-139-5p (miRNAs w/seed CUACAGU)	4.04E-03	DDIT4,HOXA9,NDFIP1,NFIA,SOCS2
PTGS1	4.08E-03	PTGER4,PTGS2
HBEGF	4.08E-03	EREG,PTGS2
diclofenac	4.08E-03	MPO,PTGS2
fulvestrant	4.14E-03	BGLAP,DDIT4,GUCY1B3,PTGS2
indomethacin	4.47E-03	CD34,MPO,PTGER4,PTGS2
Hsp27	4.57E-03	BGLAP,PTGS2
IFNE	4.57E-03	PTGER4,PTGS2
TGFB1	4.58E-03	ALDH2,BGLAP,CD34,DDIT4,EREG,LTBP3,MPP6,PTGER4,PTGS2,RBPMS,ROBO3
ATF4	4.65E-03	BGLAP,DDIT4,PTGS2
NFE2L2	4.82E-03	BGLAP,CBR1,GSTM5,MAFF,PTGS2
miR-1277-3p (and other miRNAs w/seed ACGUAGA)	4.82E-03	CPNE3,NDFIP1
IKZF1	4.87E-03	BLNK,DOCK1,SH3BP5
IKBKB	4.88E-03	EREG,LTB,PTGS2,SOCS2

ZBTB16	5.08E-03	BGLAP,CD34
AGTR1	5.08E-03	GSTM5,PTGS2
TXNIP	5.08E-03	DDIT4,PTGS2
H89	5.10E-03	BGLAP,EREG,PTGS2
IL27	5.21E-03	C10orf10,IL12RB2,PTGS2
miR-155-5p (miRNAs w/seed UAAUGCU)	5.31E-03	HOOK1,LPAR6,NDFIP1,NFIA,RAB34,ZFP3
STK11	5.33E-03	EREG,LTB,PTGS2
Growth hormone	5.45E-03	BGLAP,GSTM5,SOCS2
arbutin	5.59E-03	PTGS2
1-alpha,24(R),25-trihydroxyvitamin D3	5.59E-03	BGLAP
prostaglandin E3	5.59E-03	PTGS2
24R,25-dihydroxyvitamin D3	5.59E-03	BGLAP
des-Arg(10)-kallidin	5.59E-03	PTGS2
sPla2	5.59E-03	PTGS2
cyanidin 3-O-glucoside	5.59E-03	PTGS2
ASB2	5.59E-03	HOXA9
VRK2	5.59E-03	PTGS2
NOX5	5.59E-03	PTGS2

TMEM8B	5.59E-03	PTGS2
soy isoflavones	5.59E-03	PTGS2
3,4-dihydroxyphenylethanol	5.59E-03	PTGS2
PTPRG	5.59E-03	CD34
PLA2G2F	5.59E-03	PTGS2
IDE	5.59E-03	SNCA
OSTF1	5.59E-03	BGLAP
ECSIT	5.59E-03	PTGS2
CBR1	5.59E-03	PTGS2
ARHGDIB	5.59E-03	PTGS2
SH3GLB2	5.59E-03	PTGS2
SC68376	5.59E-03	PTGS2
pyridoxal	5.59E-03	PTGS2
zileuton	5.59E-03	PTGS2
bifenthrin	5.59E-03	PTGS2
thiamin pyrophosphate	5.59E-03	MPO
benzylamine	5.59E-03	PTGS2
bumetanide	5.59E-03	PTGS2
incyclinide	5.59E-03	PTGS2

methylprednisolone acetate	5.59E-03	PTGS2
vanillic acid	5.59E-03	PTGS2
indolo(3,2-b)carbazole	5.59E-03	SOCS2
tetrahydropalmatine	5.59E-03	PTGS2
epoxyeicosatrienoic acid	5.59E-03	PTGS2
poly(ADP-ribose)	5.59E-03	PTGS2
laminaran	5.59E-03	PTGS2
lipooligosaccharide	5.59E-03	PTGS2
1beta,25-dihydroxyvitamin D3	5.59E-03	BGLAP
BAPTA-AM	5.61E-03	PTGS2,SOCS2
ICAM1	5.89E-03	IL12RB2,MPO
miR-3680-5p (miRNAs w/seed ACUCACU)	5.89E-03	HOOK1,RBPMS
miR-4716-5p (miRNAs w/seed CCAUGUU)	5.89E-03	NFIA,OSBPL3
nonylphenol	5.89E-03	EREG,PTGS2
IL18	6.32E-03	IL12RB2,MPO,PTGS2
ARHGAP21	6.46E-03	PTGS2,SH3BP5
N-acetylsphingosine	6.46E-03	GSTM5,PTGS2

AGT	6.68E-03	EREG,GSTM5,HOXA9,PTGS2,SNCA
luteolin	6.76E-03	MPO,PTGS2
WT1	6.99E-03	EREG,SH3BP5,SLC35G2
PPRC1	7.06E-03	DDIT4,PTGS2
APOA1	7.06E-03	MPO,PTGS2
VitaminD3-VDR-RXR	7.37E-03	BGLAP,HOXA10
EGFR	7.49E-03	EREG,HOXA7,PTGER4,PTGS2
baicalin	7.68E-03	BGLAP,CBR1
DUSP1	7.68E-03	IL12RB2,PTGS2
mifepristone	7.97E-03	DDIT4,EPCAM,HOXA10,PTGS2
EPO	8.07E-03	CD52,MPO,PTGS2,SOCS2
prostaglandin E2	8.27E-03	EREG,PTGER4,PTGS2,SOCS2
NOG	8.33E-03	BGLAP,PTGS2
DKK1	8.33E-03	BGLAP,EPCAM
sulprostone	8.37E-03	PTGS2
chloride	8.37E-03	PTGS2
W146	8.37E-03	PTGS2
imperatorin	8.37E-03	PTGS2
Eif2	8.37E-03	PTGS2
pycnogenols	8.37E-03	PTGS2

pinoresinol	8.37E-03	PTGS2
HN	8.37E-03	SH3BP5
RGD1560225	8.37E-03	PTGS2
SPEN	8.37E-03	BGLAP
CTSK	8.37E-03	PTGS2
ERC1	8.37E-03	PTGS2
AA-861	8.37E-03	PTGS2
amifostine	8.37E-03	PTGS2
trimetazidine	8.37E-03	MPO
tenidap	8.37E-03	PTGS2
acacetin	8.37E-03	PTGS2
flavone	8.37E-03	PTGS2
12-hydroxyeicosatetraenoic acid	8.37E-03	PTGS2
epiallopregnanolone	8.37E-03	PTGS2
IRF1	8.61E-03	IL12RB2,LTB,PTGS2
ZBTB20	8.66E-03	GSTM5,SOCS2
miR-517a-3p (and other miRNAs w/seed UCGUGCA)	8.66E-03	HOXA5,NFIA
miR-124-3p (and other miRNAs w/seed AAGGCAC)	8.98E-03	ATP6V0A2,CPNE3,ERMP1,HOXA5,NAP1L3,NDFIP1,NFIA,OSBPL3,RAB34,RHBDF1,SLITRK5

Gm-csf	9.00E-03	PTGS2,SOCS2
PDGFB	9.00E-03	BGLAP,PTGS2
FGF10	9.00E-03	PTGS2,VNN1
glutathione	9.00E-03	MPO,PTGS2
miR-1321 (and other miRNAs w/seed AGGGAGG)	9.20E-03	CD34,CHRDL1,HOXA10,OSBPL3
bucladesine	9.30E-03	BGLAP,FAAH,PTGER4,PTGS2
carbonyl cyanide m-chlorophenyl hydrazone	9.34E-03	DDIT4,PTGS2
2-deoxyglucose	9.69E-03	DDIT4,PTGS2
SRF	9.74E-03	GSTM5,MPPED2,MS4A3,PTGS2
zymosan	1.00E-02	C10orf10,PTGS2
BNIP3L	1.04E-02	BLNK,LTB
hemin	1.08E-02	GUCY1B3,PTGS2
miR-4793-5p (miRNAs w/seed CAUCCUG)	1.11E-02	ALDH2,NDFIP1
HDAC3	1.11E-02	CD34,PTGS2
Nfatc	1.11E-02	PTGS2
8-chloro-cAMP	1.11E-02	PTGS2
2-(3-hydroxypropoxy)calcitriol	1.11E-02	BGLAP

NF-kappaB decoy	1.11E-02	PTGS2
trans-cinnamaldehyde	1.11E-02	PTGS2
tylophorine	1.11E-02	PTGS2
nebivolol	1.11E-02	PTGS2
RSPO3	1.11E-02	PTGS2
LEO1	1.11E-02	HOXA9
SGPL1	1.11E-02	PTGS2
ARID2	1.11E-02	BGLAP
LIN9	1.11E-02	BGLAP
TMEM119	1.11E-02	BGLAP
betulin	1.11E-02	PTGS2
PPP1R1B	1.11E-02	SNCA
ENPP1	1.11E-02	BGLAP
MUC2	1.11E-02	PTGS2
HSD11B2	1.11E-02	BGLAP
PHF19	1.11E-02	HOXA5
NUCB2	1.11E-02	PTGS2
TDO2	1.11E-02	PTGS2
TIA1	1.11E-02	PTGS2
UBE2D1	1.11E-02	PTGS2
AJUBA	1.11E-02	DOCK1

SKLB023	1.11E-02	PTGS2
farnesyl transferase	1.11E-02	MPO
SULT1E1	1.11E-02	PTGS2
ANXA6	1.11E-02	BGLAP
HDAC8	1.11E-02	HOXA5
furosemide	1.11E-02	PTGS2
1-butanol	1.11E-02	PTGS2
lestaurtinib	1.11E-02	MPO
lansoprazole	1.11E-02	PTGS2
7,8-dihydro-7,8-dihydroxybenzo(a)pyrene 9,10-oxide	1.11E-02	PTGS2
ethyl linoleate	1.11E-02	PTGS2
cadmium sulfate	1.11E-02	PTGS2
polyphosphate	1.11E-02	BGLAP
vanadium pentoxide	1.11E-02	PTGS2
aclarubicin	1.11E-02	PTGS2
ecdysterone	1.11E-02	BGLAP
FOXL2	1.15E-02	MAFF,PTGS2
NFATC3	1.15E-02	DDIT4,PTGS2
miR-3944-5p (miRNAs w/seed GUGCAGC)	1.17E-02	CD34,NFIA,SH3BP5

miR-3974 (miRNAs w/seed AAGGUCA)	1.19E-02	ERMP1,MS4A3
seocalcitol	1.19E-02	BGLAP,PTGS2
Salmonella enterica serotype abortus equi lipopolysaccharide	1.21E-02	EREG,MAFF,PTGS2
PPARD	1.21E-02	ALDH2,PTGER4,PTGS2
8-bromo-cAMP	1.23E-02	ALDH2,BGLAP,PTGS2
PTHLH	1.23E-02	BGLAP,PTGS2
PDGF BB	1.24E-02	EREG,GUCY1B3,LAMB2,PTGS2
SP1	1.25E-02	BGLAP,EREG,IL12RB2,LTB,PTGS2
INHBA	1.27E-02	CPNE8,EREG,MPPED2
miR-3649 (miRNAs w/seed GGGACCU)	1.29E-02	CPNE3,HOXA10,MAFF
bortezomib	1.29E-02	BGLAP,PTGS2,SNCA
MAPK1	1.31E-02	BGLAP,OSBPL3,PTGS2,SOCS2
IRF2	1.31E-02	IL12RB2,PTGS2
miR-3115 (miRNAs w/seed UAUGGGU)	1.31E-02	HDAC9,PTGER4
miR-3928-3p (miRNAs w/seed GAGGAAC)	1.31E-02	GSTM5,OSBPL3
CD14	1.31E-02	MPO,PTGS2
INSR	1.33E-02	ALDH2,CBR1,EREG,SOCS2

STAT5A	1.35E-02	IL12RB2,LTB,SOCS2
15-deoxy-delta-12,14 -PGJ 2	1.37E-02	BGLAP,MPO,PTGS2
MAP3K14	1.39E-02	LTB,PTGS2
ILX-23-7553	1.39E-02	BGLAP
pyrophosphate	1.39E-02	BGLAP
nitrate	1.39E-02	MPO
dibutyryl cGMP	1.39E-02	PTGS2
dienogest	1.39E-02	PTGS2
Cpla2	1.39E-02	PTGS2
TET1	1.39E-02	HOXA9
elocalcitol	1.39E-02	BGLAP
GPR68	1.39E-02	PTGS2
carvacrol	1.39E-02	PTGS2
PARG	1.39E-02	PTGS2
Atf	1.39E-02	PTGS2
PLCE1	1.39E-02	PTGS2
CDH4	1.39E-02	PTGS2
PAPOLA	1.39E-02	PTGS2
DSPP	1.39E-02	BGLAP
FIGF	1.39E-02	BGLAP

PLAA	1.39E-02	PTGS2
2-cyclohexen-1-one	1.39E-02	PTGS2
pyridoxal phosphate-6-azophenyl-2',4'-disulfonic acid	1.39E-02	PTGS2
mevalonolactone	1.39E-02	PTGS2
3-aminotriazole	1.39E-02	MPO
butylated hydroxytoluene	1.39E-02	PTGS2
NADPH	1.39E-02	PTGS2
5-hydroxytryptophan	1.39E-02	PTGS2
N-acetylglucosamine	1.39E-02	PTGS2
ginsenoside Rg1	1.39E-02	PTGS2
miR-144-3p (miRNAs w/seed ACAGUAU)	1.40E-02	CLIC2,HOXA10,HOXA7,NDFIP1,NFIA,PTGS2,SOCS2
progesterone	1.42E-02	FAAH,HOXA10,MPO,PTGER4,PTGS2
POR	1.43E-02	EREG,GSTM5,VNN1
IL17A	1.43E-02	EREG,MPO,PTGS2
miR-2355-3p (miRNAs w/seed UUGUCCU)	1.43E-02	ATP6V0A2,ERMP1
TBP	1.43E-02	HOXA9,PTGS2
ATP-gamma-S	1.52E-02	EREG,PTGS2

RARB	1.52E-02	HOXA5,PTGS2
CDX2	1.52E-02	HOXA5,HOXA9
SPIB	1.52E-02	BLNK,EPCAM
cobalt chloride	1.52E-02	DDIT4,PTGS2
L-triiodothyronine	1.52E-02	BGLAP,CBR1,PTGS2,RAB34
estradiol benzoate	1.56E-02	PTGER4,PTGS2
1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine	1.56E-02	MPO,PTGS2
IL4	1.58E-02	ALDH2,FAAH,IL12RB2,LTB,PTGS2,SOCS2
DOCK8	1.60E-02	PTGS2,SH3BP5
SPDEF	1.60E-02	CD34,LAMB2
MAP2K3	1.60E-02	PTGS2,RAB34
miR-448-3p (and other miRNAs w/seed UGCAUAU)	1.61E-02	DDIT4,MPPED2,NFIA,SLITRK5,SOCS2,ZFP3
miR-223-3p (miRNAs w/seed GUCAGUU)	1.64E-02	DDIT4,NFIA,SLC35G2,SNCA
EDN1	1.65E-02	EREG,PTGER4,PTGS2
miR-1839-5p (and other miRNAs w/seed AGGUAGA)	1.65E-02	HDAC9,HOXA7
nitrite	1.67E-02	MPO
teichoic acid	1.67E-02	PTGS2

LRRC26	1.67E-02	LTB
CACNA1C	1.67E-02	PTGS2
EN2	1.67E-02	SNCA
PDGFD	1.67E-02	PTGS2
PTGDR	1.67E-02	PTGS2
SSPN	1.67E-02	PTGS2
DES	1.67E-02	HOXA10
OGN	1.67E-02	BGLAP
TAB2	1.67E-02	PTGS2
PROK1	1.67E-02	PTGS2
PIM3	1.67E-02	IL12RB2
WAC	1.67E-02	BGLAP
glaucocalyxin A	1.67E-02	PTGS2
RNF17	1.67E-02	MPO
SNW1	1.67E-02	BGLAP
XPA	1.67E-02	PTGS2
PITPNA	1.67E-02	PTGS2
Pde4d	1.67E-02	PTGS2
ambroxol	1.67E-02	MPO
EGTA acetoxymethyl ester	1.67E-02	PTGS2

olmesartan medoxomil	1.67E-02	PTGS2
GW 5074	1.67E-02	PTGS2
4-nitroquinoline-1-oxide	1.67E-02	PTGS2
ferric nitrilotriacetate	1.67E-02	PTGS2
RWJ 67657	1.67E-02	PTGS2
dibenzoylmethane	1.67E-02	PTGS2
ibandronic acid	1.67E-02	BGLAP
dieldrin	1.67E-02	SNCA
propyl gallate	1.67E-02	PTGS2
phosphatidic acid	1.67E-02	PTGS2
phenylamil	1.67E-02	BGLAP
canrenoate potassium	1.67E-02	PTGS2
miR-3615 (miRNAs w/seed CUCUCGG)	1.69E-02	CD34,SLITRK5
SASH1	1.69E-02	PTGS2,SH3BP5
ERG	1.69E-02	DOCK1,GUCY1B3,SNCA
SMAD4	1.76E-02	BGLAP,EREG,PTGS2
isotretinoin	1.83E-02	HOXA5,PTGS2
SP600125	1.86E-02	BGLAP,DDIT4,PTGS2
TRAF2	1.88E-02	EPCAM,RHBDF1

MET	1.92E-02	PTGS2,SH3BP5
GW9662	1.92E-02	MPO,PTGS2
ascorbic acid	1.92E-02	BGLAP,PTGS2
16,16-dimethylprostaglandin E2	1.94E-02	PTGER4
(6)-gingerol	1.94E-02	PTGS2
Rhox4b (includes others)	1.94E-02	CD34
Glucocorticoid-GCR	1.94E-02	BGLAP
CTR9	1.94E-02	HOXA9
sphingomyelinase	1.94E-02	PTGS2
DVL1	1.94E-02	PTGS2
RAPGEF1	1.94E-02	PTGS2
ENO1	1.94E-02	PTGS2
TPSD1	1.94E-02	PTGS2
LPAR2	1.94E-02	PTGS2
APOBEC1	1.94E-02	PTGS2
SF3B2	1.94E-02	HOXA5
MLLT3	1.94E-02	HOXA9
IL18RAP	1.94E-02	PTGS2
PLA2G2A	1.94E-02	PTGS2

NEDD4	1.94E-02	SNCA
4-coumaric acid	1.94E-02	PTGS2
zimelidine	1.94E-02	C10orf10
eugenol	1.94E-02	PTGS2
atropine	1.94E-02	PTGS2
S-allyl-L-cysteine	1.94E-02	PTGS2
20alpha-hydroxycholesterol	1.94E-02	BGLAP
IL13	1.97E-02	CD52,PTGS2,SNCA,VNN1
TGFA	1.97E-02	EREG,PTGS2
miR-3180-3p (and other miRNAs w/seed GGGGCGG)	1.99E-02	DDIT4,NFIA,SLITRK5,ZFP3
GATA2	2.02E-02	CD34,MPO
miR-3691-5p (miRNAs w/seed GUGGAUG)	2.02E-02	CPNE8,CRHBP
FGFR1	2.02E-02	BGLAP,PTGS2
SNCA	2.02E-02	BGLAP,SNCA
imatinib	2.02E-02	BLNK,SOCS2
ESR2	2.03E-02	EREG,PTGS2,SOCS2
CEBPB	2.06E-02	BGLAP,BLNK,PTGER4,PTGS2
TRAF3	2.07E-02	EPCAM,RHBDF1

actinomycin D	2.08E-02	GSTM5,GUCY1B3,PTGS2
ciglitazone	2.12E-02	MPO,PTGS2
miR-4318 (miRNAs w/seed ACUGUGG)	2.17E-02	EPCAM,MREG
6-hydroxydopamine	2.17E-02	DDIT4,SNCA
eicosapentenoic acid	2.17E-02	PTGS2,SNCA
2',3'-dialdehyde ATP	2.22E-02	PTGS2
1L-6-hydroxymethyl-chiro-inositol 2-(R)-2-O-methyl-3-O-octadecylcarbonate	2.22E-02	PTGS2
ganglioside GD1a	2.22E-02	PTGS2
ganglioside GM1	2.22E-02	PTGS2
desoxycorticosterone	2.22E-02	PTGS2
SMTNL1	2.22E-02	PTGS2
atrazine	2.22E-02	PTGS2
voltage-gated calcium channel	2.22E-02	CD34
Cyclin A	2.22E-02	CD34
BCG vaccine	2.22E-02	PTGS2
TCF/LEF	2.22E-02	PTGS2
RNF40	2.22E-02	BGLAP
EN1	2.22E-02	SNCA

TFRC	2.22E-02	SNCA
IL1R2	2.22E-02	MPO
GRIP1	2.22E-02	PTGS2
CBX8	2.22E-02	HOXA9
VEGFC	2.22E-02	PTGS2
PIM2	2.22E-02	IL12RB2
DLX5	2.22E-02	BGLAP
S1PR3	2.22E-02	PTGS2
PLA2G5	2.22E-02	PTGS2
WNT3	2.22E-02	PTGS2
naproxen	2.22E-02	PTGS2
allyl isothiocyanate	2.22E-02	PTGS2
NCX-4040	2.22E-02	PTGS2
capsazepine	2.22E-02	PTGS2
flufenamic acid	2.22E-02	PTGS2
lapatinib	2.22E-02	EREG
carnosol	2.22E-02	PTGS2
glimepiride	2.22E-02	BGLAP
acetic acid	2.22E-02	MPO
palmitoyl-Cys((RS)-2,3-di(palmitoyloxy)-propyl)-Ala-Gly-OH	2.22E-02	PTGS2

rutin	2.22E-02	PTGS2
tosyllysine chloromethyl ketone	2.22E-02	PTGS2
mannan	2.22E-02	PTGS2
22(S)-hydroxycholesterol	2.22E-02	BGLAP
lead acetate	2.22E-02	GSTM5
MYB	2.22E-02	CD34,PTGS2
TBK1	2.22E-02	PTGS2,SH3BP5
MKL2	2.27E-02	GSTM5,MS4A3
miR-3622a-3p (and other miRNAs w/seed CACCUGA)	2.27E-02	ATP6V0A2,FAM169A
allopurinol	2.32E-02	MPO,PTGS2
miR-4438 (miRNAs w/seed ACAGGCU)	2.32E-02	HOXA5,SH3BP5
FBXO32	2.32E-02	HOOK1,PTGS2
VEGFA	2.33E-02	ALDH2,CD34,PTGS2
WNT3A	2.33E-02	BGLAP,DDIT4,PTGS2
okadaic acid	2.37E-02	BGLAP,PTGS2
TFAP2A	2.42E-02	EREG,PTGER4
Rxr	2.48E-02	BGLAP,PTGS2
NRAS	2.48E-02	CD34,PTGS2

daidzein	2.48E-02	OSBPL3,PTGS2
TCF	2.48E-02	EPCAM,PTGS2
ganglioside	2.49E-02	PTGS2
tannic acid	2.49E-02	PTGS2
ganglioside GT1	2.49E-02	PTGS2
RUNX1T1	2.49E-02	CD34
dovitinib	2.49E-02	LTB
FAT1	2.49E-02	PTGS2
NEK7	2.49E-02	PTGS2
SLC18A2	2.49E-02	SNCA
SOX8	2.49E-02	VNN1
mir-194	2.49E-02	SOCS2
CD209	2.49E-02	PTGS2
PRDX1	2.49E-02	PTGS2
NEK6	2.49E-02	PTGS2
TAF6	2.49E-02	HOXA9
BTC	2.49E-02	PTGS2
HPSE	2.49E-02	PTGS2
ammonium trichloro(dioxoethylene-O,O'-)tellurate	2.49E-02	MPO
bupivacaine	2.49E-02	MPO

domoic acid	2.49E-02	PTGS2
phenyl-N-tert-butylnitrone	2.49E-02	PTGS2
racemic flurbiprofen	2.49E-02	PTGS2
ketorolac	2.49E-02	PTGS2
mesalamine	2.49E-02	PTGS2
rhodioloside	2.49E-02	BGLAP
nickel chloride	2.49E-02	PTGS2
hexamethoxyflavone	2.49E-02	PTGS2
theophylline	2.49E-02	MPO
tyrphostin AG 127	2.49E-02	PTGS2
tyrphostin AG 1024	2.49E-02	PTGS2
saturated fatty acid	2.49E-02	PTGS2
arachidonyltrifluoromethane	2.49E-02	PTGS2
L-N6-(1-iminoethyl)-lysine	2.49E-02	PTGS2
miR-342-5p (and other miRNAs w/seed GGGGUGC)	2.49E-02	CD34,LST1,NFIA,ZFP3
PPARG	2.49E-02	BGLAP,MPO,PTGS2,VNN1
PARP1	2.53E-02	PTGS2,SOCS2
fatty acid	2.53E-02	PTGS2,VNN1

MYD88	2.58E-02	MAFF,PTGS2,SH3BP5
TAL1	2.58E-02	DOCK1,MPO
phorbol esters	2.58E-02	HOXA5,PTGS2
camptothecin	2.59E-02	LAMB2,LST1,MAFF,PTGER4,PTGS2
PPARA	2.64E-02	ALDH2,PTGS2,SOCS2,VNN1
SAMSN1	2.69E-02	PTGS2,SH3BP5
aspirin	2.69E-02	MPO,PTGS2
EPHB1	2.76E-02	PTGS2
(+)-catechin	2.76E-02	PTGS2
Pdgfr	2.76E-02	EREG
FFAR4	2.76E-02	PTGS2
SGPP1	2.76E-02	PTGS2
NQO2	2.76E-02	PTGS2
FOXN1	2.76E-02	MREG
mir-130	2.76E-02	HOXA5
NMNAT1	2.76E-02	SOCS2
RYBP	2.76E-02	HOXA7
PTAFR	2.76E-02	PTGS2
VIM	2.76E-02	BGLAP
DEFB103A/DEFB103B	2.76E-02	PTGS2

TAS1R3	2.76E-02	DDIT4
edaravone	2.76E-02	PTGS2
methylene blue	2.76E-02	MPO
clomipramine	2.76E-02	C10orf10
cinnamaldehyde	2.76E-02	PTGS2
rosmarinic acid	2.76E-02	PTGS2
morin	2.76E-02	PTGS2
midostaurin	2.76E-02	PTGS2
8-oxo-7-hydrodeoxyguanosine	2.76E-02	PTGS2
allopregnanolone	2.76E-02	PTGS2
formononetin	2.76E-02	BGLAP
miR-502-5p (and other miRNAs w/seed UCCUUGC)	2.80E-02	MARC2,MREG
salmonella minnesota R595 lipopolysaccharides	2.86E-02	MAFF,PTGS2
bisindolylmaleimide I	2.91E-02	BGLAP,PTGS2
tamoxifen	2.95E-02	C10orf10,MAFF,PTGS2
ERBB2	2.95E-02	CD34,EREG,LTBP3,PTGS2,RAB34
miR-224-5p (miRNAs w/seed AAGUCAC)	2.96E-02	CPNE8,HOXA5,MAFF,NFIA

miR-128-3p (and other miRNAs w/seed CACAGUG)	2.98E-02	HOXA10,HOXA5,HOXA9,LPAR6,MPPED2,NFIA,RAB34
TRIM24	3.03E-02	BLNK,SOCS2
L-histidine	3.04E-02	DDIT4
estriol	3.04E-02	HOXA10
DNA-methyltransferase	3.04E-02	PTGS2
CARD9	3.04E-02	MPO
Agtr1b	3.04E-02	PTGS2
SYK/ZAP	3.04E-02	PTGS2
PP2A	3.04E-02	PTGS2
Rac	3.04E-02	PTGS2
trypsin	3.04E-02	PTGS2
Adaptor protein 1	3.04E-02	PTGS2
T 0070907	3.04E-02	PTGS2
Collagen Alpha1	3.04E-02	PTGS2
IL13RA2	3.04E-02	VNN1
PLD1	3.04E-02	PTGS2
MTA3	3.04E-02	BLNK
MAPK10	3.04E-02	PTGS2
ITGA3	3.04E-02	PTGS2

MZF1	3.04E-02	CD34
HNRNPAB	3.04E-02	PTGS2
sesamin	3.04E-02	PTGS2
manumycin A	3.04E-02	PTGS2
amiloride	3.04E-02	PTGS2
enterolactone	3.04E-02	BGLAP
phospholipid	3.04E-02	PTGS2
puerarin	3.04E-02	PTGS2
tyrphostin AG 1296	3.04E-02	PTGS2
miR-200b-3p (and other miRNAs w/seed AAUACUG)	3.05E-02	ATP6V0A2,CHRDL1,CRHBP,DDIT4,HOOK1,HOXA5,NFIA
fenofibrate	3.07E-02	HOXA7,MPO,PTGS2
Ca2+	3.07E-02	BGLAP,FAAH,PTGS2
GNA15	3.08E-02	GUCY1B3,LAMB2
GW501516	3.14E-02	PTGER4,VNN1
hyaluronic acid	3.20E-02	MPO,PTGS2
CD3	3.20E-02	GUCY1B3,IL12RB2,PTGER4,PTGS2,SOCS2
hydrogen peroxide	3.22E-02	BGLAP,DDIT4,PTGER4,PTGS2
NFKBIA	3.22E-02	EREG,GSTM5,HOXA10,PTGS2
MAPT	3.23E-02	ERMP1,GSTM5,SNCA
F2	3.23E-02	EREG,HDAC9,PTGS2

cephaloridine	3.26E-02	DDIT4,GSTM5
bisphenol A	3.26E-02	EREG,PTGS2
miR-3189-5p (miRNAs w/seed GCCCCAU)	3.27E-02	FAM169A,MPPED2,SLITRK5
PD173074	3.31E-02	LTB
NFkB (family)	3.31E-02	PTGS2
perhexiline	3.31E-02	C10orf10
chlorcyclizine	3.31E-02	C10orf10
HRH1	3.31E-02	PTGS2
STAT1/3/5 dimer	3.31E-02	SOCS2
harmine	3.31E-02	PTGS2
astragalin	3.31E-02	PTGS2
SLAMF1	3.31E-02	IL12RB2
PLA2G10	3.31E-02	PTGS2
MIR101	3.31E-02	PTGS2
mir-214	3.31E-02	PTGS2
miR-643 (miRNAs w/seed CUUGUAU)	3.31E-02	EREG
KLF10	3.31E-02	BGLAP
TRPC1	3.31E-02	SNCA
NPPC	3.31E-02	GUCY1B3

10-nitrooleate	3.31E-02	PTGS2
GNA13	3.31E-02	PTGS2
PDCD4	3.31E-02	PTGS2
RHOB	3.31E-02	PTGS2
ebselen	3.31E-02	PTGS2
N-(3-oxododecanoyl)-homoserine lactone	3.31E-02	PTGS2
dimethyl fumarate	3.31E-02	PTGS2
D-sphingosine	3.31E-02	PTGS2
SKF-38393	3.31E-02	BGLAP
N,N-dimethylsphingosine	3.31E-02	PTGS2
buthionine sulfoximine	3.31E-02	PTGS2
progestin	3.31E-02	HOXA10
NR1I3	3.32E-02	GSTM5,MAFF
KAT5	3.32E-02	EREG,HOXA9
NR3C1	3.38E-02	BGLAP,DDIT4,LTB,PTGS2,SNCA
INHA	3.38E-02	EREG,PTGS2
cyclic AMP	3.40E-02	BGLAP,EPCAM,PTGS2
miR-4687-3p (miRNAs w/seed GGCUGUU)	3.44E-02	C10orf10,HDAC9

JAK2	3.44E-02	MPO,PTGS2
RB1	3.50E-02	BGLAP,H1F0,ROBO3
miR-361-5p (miRNAs w/seed UAUCAGA)	3.54E-02	ATP6V0A2,GUCY1B3,NFIA
MKL1	3.56E-02	GSTM5,MS4A3
RARA	3.56E-02	HOXA5,PTGS2
leukotriene C4	3.58E-02	PTGS2
SIM2	3.58E-02	ROBO3
TRPV1	3.58E-02	PTGS2
LPAR1	3.58E-02	PTGS2
CSF3R	3.58E-02	MPO
CD46	3.58E-02	PTGER4
PTGFR	3.58E-02	PTGS2
PTGER1	3.58E-02	PTGS2
PLD2	3.58E-02	PTGS2
SP7	3.58E-02	BGLAP
FABP1	3.58E-02	FAAH
Go6983	3.58E-02	PTGS2
1,2-dimethylhydrazine	3.58E-02	PTGS2
mevastatin	3.58E-02	PTGS2

SU5402	3.58E-02	LTB
icatibant	3.58E-02	PTGS2
farnesyl pyrophosphate	3.58E-02	PTGS2
gamma-linolenic acid	3.58E-02	PTGS2
prostaglandin	3.58E-02	PTGS2
auranofin	3.58E-02	PTGS2
IL5	3.68E-02	LTB,RBPMS,SOCS2
CCND1	3.75E-02	CPNE3,EREG,MAFF
miR-381-3p (and other miRNAs w/seed AUACAAG)	3.76E-02	ATP6V0A2,HOXA9,NDFIP1,NFIA,OSBPL3,SNCA
miR-3198 (and other miRNAs w/seed UGGAGUC)	3.81E-02	IL12RB2,SCRN1
miR-4421 (and other miRNAs w/seed CCUGUCU)	3.81E-02	H1F0,MANSC1
HDAC2	3.81E-02	CD34,PTGS2
A23187	3.81E-02	DDIT4,PTGS2
sulforafan	3.81E-02	GSTM5,PTGS2
prostaglandin A2	3.85E-02	LPAR6
lonafarnib	3.85E-02	PTGS2
cyclooxygenase	3.85E-02	PTGS2

IKK (complex)	3.85E-02	PTGS2
dihydroartemisinin	3.85E-02	PTGS2
KLF9	3.85E-02	HOXA10
mir-135	3.85E-02	HOXA10
PDGFRA	3.85E-02	BGLAP
PRDX2	3.85E-02	PTGS2
BAG1	3.85E-02	PTGS2
salmeterol	3.85E-02	DDIT4
hydrogen sulfide	3.85E-02	PTGS2
N1,N11-diethylnorspermine	3.85E-02	PTGS2
myricetin	3.85E-02	PTGS2
neuroprotectin D1	3.85E-02	PTGS2
gambogic acid	3.85E-02	PTGS2
Cd2+	3.85E-02	PTGS2
manganese	3.85E-02	PTGS2
paraquat	3.93E-02	PTGS2,SNCA
miR-26a-5p (and other miRNAs w/seed UCAAGUA)	3.96E-02	ERMP1,HOOK1,HOXA5,HOXA9,MREG,PTGS2
corticosterone	4.06E-02	DDIT4,PTGS2
IL21	4.06E-02	IL12RB2,SOCS2

miR-3103-5p (and other miRNAs w/seed GAGGGAG)	4.11E-02	ALDH2,CD34,CHRDL1
epoprostenol	4.12E-02	PTGS2
cyclic GMP	4.12E-02	IL12RB2
TMSB4	4.12E-02	BGLAP
1'-acetoxychavicol acetate	4.12E-02	PTGS2
IGFBP7	4.12E-02	PTGS2
MAPK8IP1	4.12E-02	SNCA
PLCG2	4.12E-02	PTGS2
CD83	4.12E-02	PTGS2
S1PR2	4.12E-02	PTGS2
SP2	4.12E-02	BGLAP
RPS6KA5	4.12E-02	PTGS2
ITGAL	4.12E-02	IL12RB2
RND3	4.12E-02	PTGS2
CD47	4.12E-02	IL12RB2
pyridoxine	4.12E-02	PTGS2
rofecoxib	4.12E-02	PTGS2
caffeic acid	4.12E-02	PTGS2
lauric acid	4.12E-02	PTGS2

resveratrol	4.27E-02	BGLAP,MPO,PTGS2
miR-4300 (and other miRNAs w/seed GGGAGCU)	4.30E-02	GSTM5,HOXA7,MPPED2
POU2F1	4.32E-02	DDIT4,GSTM5
lovastatin	4.32E-02	MPP6,PTGS2
misoprostol	4.38E-02	PTGS2
RNASEL	4.38E-02	PTGS2
SALL4	4.38E-02	EPCAM
CGB (includes others)	4.38E-02	PTGS2
PRKCI	4.38E-02	PTGS2
HDC	4.38E-02	PTGS2
SCGB1A1	4.38E-02	PTGS2
superoxide	4.38E-02	PTGS2
allyl sulfide	4.38E-02	GSTM5
D609	4.38E-02	PTGS2
amlodipine	4.38E-02	PTGS2
piperine	4.38E-02	PTGS2
2-arachidonoylglycerol	4.38E-02	PTGS2
leptomycin B	4.38E-02	PTGS2
acadesine	4.38E-02	PTGS2

Alpha catenin	4.39E-02	EREG,PTGS2
BMP7	4.39E-02	BGLAP,CD34
miR-340-5p (miRNAs w/seed UAUAAAG)	4.41E-02	CPNE8,GUCY1B3,H1F0,HOXA10,MAFF,MPPED2,NFIA,RBPMS
IL1RN	4.52E-02	HDAC9,PTGS2
miR-324-5p (miRNAs w/seed GCAUCCC)	4.58E-02	HOXA9,MPPED2,ZFP3
interferon beta-1a	4.59E-02	LTB,SNCA
5-azacytidine	4.59E-02	MAFF,PTGS2
miR-130a-3p (and other miRNAs w/seed AGUGCAA)	4.62E-02	EREG,HOXA5,MPPED2,NAP1L3,NFIA,RAB34
leukotriene B4	4.65E-02	MPO
calcipotriene	4.65E-02	BGLAP
Cyclin E	4.65E-02	CD34
NMDA Receptor	4.65E-02	PTGS2
HCAR2	4.65E-02	PTGS2
WWTR1	4.65E-02	BGLAP
TEAD1	4.65E-02	PTGS2
Endothelin	4.65E-02	PTGS2
fontolizumab	4.65E-02	IL12RB2
p85 (pik3r)	4.65E-02	PTGS2

CBX4	4.65E-02	HOXA7
FKBP4	4.65E-02	HOXA10
miR-4766-5p (and other miRNAs w/seed CUGAAAG)	4.65E-02	ATP6V0A2
WNT7A	4.65E-02	HOXA10
TACR1	4.65E-02	PTGS2
S1PR1	4.65E-02	PTGS2
PD 169316	4.65E-02	PTGS2
ellagic acid	4.65E-02	PTGS2
ethoxyquin	4.65E-02	GSTM5
NO 1886	4.65E-02	PTGS2
ethyl pyruvate	4.65E-02	PTGS2
dipyridamole	4.65E-02	PTGS2
isoliquiritigenin	4.65E-02	PTGS2
magnolol	4.65E-02	PTGS2
ammonium chloride	4.65E-02	SNCA
vanadate	4.65E-02	BGLAP
miR-335-5p (and other miRNAs w/seed CAAGAGC)	4.65E-02	NDFIP1,OSBPL3,SCRN1
SMAD7	4.73E-02	HDAC9,LTBP3
HGF	4.75E-02	EPCAM,GUCY1B3,PTGS2,SOCS2

miR-133a-3p (and other miRNAs w/seed UUGGUCC)	4.75E-02	CPNE3,HOXA9,NFIA,RAB34,SNCA
Pdgf Ab	4.92E-02	EREG
3M-002	4.92E-02	PTGS2
artesunic acid	4.92E-02	PTGS2
miR-4652-3p (miRNAs w/seed UUCUGUU)	4.92E-02	CPNE8
HOXB9	4.92E-02	EREG
CREB3L1	4.92E-02	BGLAP
benzyl isothiocyanate	4.92E-02	PTGS2
S-(2,3-bispalmitoyloxypropyl)-cysteine-GDPKHPKSF	4.92E-02	PTGS2
IL2	4.93E-02	IL12RB2,LTB,PTGS2,SOCS2
TLR4	4.94E-02	EREG,PTGS2,SH3BP5
miR-7a-5p (and other miRNAs w/seed GGAAGAC)	4.96E-02	DDIT4,NFIA,OSBPL3,SNCA

Table 2.10. Association of LSC epigenetic signature with DMRs of genetic mutations

Columns are chromosome, start, end, diffMethyl(difference of methylation percentage for LSC-Blast), diffexp(difference of expression log2 value for LSC-Blast), Gene, DMR name, DNMT3A,IDH1,IDH2,TET2,NPM1,ASXL1(Genetic mutations tested here to look at overlap of LSC epigenetic signature and DMR for mutations), Mechanism(which mechanism regulate each DMR).

"1" represent overlap and "0" represents no overlap between the corresponding LSC epigenetic signature and DMRs for DNMT3A, IDH1, IDH2, TET2, NPM1 and ASXL1.

Gene	DMR name	DNMT3A	IDH1	IDH2	TET1	TET2	NPM1	ASXL1	Mechanism
MPO	MPO/DMR1	0	0	0	0	0	0	1	UpstremRegulator
HOXA9	HOXA9/DMR1	0	0	0	0	0	0	0	PrimaryEpi
HOXA9	HOXA9/DMR2	0	0	0	0	0	0	0	PrimaryEpi
HOXA9	HOXA9/DMR3	0	0	0	0	0	1	0	UpstremRegulator
HOXA9	HOXA9/DMR4	0	0	0	0	0	1	0	UpstremRegulator
HOXA9	HOXA9/DMR5	0	0	0	0	0	0	0	PrimaryEpi
SCRN1	SCRN1/DMR1	1	0	0	0	0	0	0	EpigeneticEnzyme
CD34	CD34	0	0	0	0	0	0	1	UpstremRegulator

CPNE3	CPNE3	0	0	0	0	0	0	0	PrimaryEpi
HOXA10	HOXA10	0	0	0	0	0	1	0	UpstremRegulator
MANSC1	MANSC1	0	0	0	0	0	0	0	PrimaryEpi
NDFIP1	NDFIP1	0	0	0	0	0	0	0	PrimaryEpi
FAM24B	FAM24B	0	0	0	1	0	0	0	EpigeneticEnzyme
FAM169A	FAM169A	1	0	0	0	0	0	0	EpigeneticEnzyme
MPPED2	MPPED2	1	0	0	0	0	0	0	EpigeneticEnzyme
PTGER4	PTGER4	0	0	0	0	0	0	0	PrimaryEpi
HOXA7	HOXA7/DMR1	0	0	0	0	0	0	0	PrimaryEpi
HOXA7	HOXA7/DMR2	1	0	0	0	0	1	0	CommonTarget
RHBDF1	RHBDF1	1	0	0	0	0	0	0	EpigeneticEnzyme
EREG	EREG	0	0	0	0	0	0	0	PrimaryEpi
CPNE8	CPNE8	0	0	0	0	0	1	0	UpstremRegulator
RAB34	RAB34/DMR1	0	0	0	0	0	0	0	PrimaryEpi
LTBP3	LTBP3	0	0	0	0	0	0	0	PrimaryEpi
RAB34	RAB34/DMR2	0	0	0	0	0	0	0	PrimaryEpi
ROBO3	ROBO3	0	0	0	0	0	0	1	UpstremRegulator
GSTM5	GSTM5	0	0	0	0	0	0	1	UpstremRegulator
DDIT4	DDIT4	0	0	0	0	0	0	1	UpstremRegulator
REC8	REC8	1	0	0	0	1	0	1	CommonTarget
ZFP3	ZFP3	0	0	0	0	0	0	0	PrimaryEpi

SLITRK5	SLITRK5	0	0	0	0	0	0	0	PrimaryEpi
HOXA6	HOXA6	1	0	0	0	0	1	0	CommonTarget
FAAH	FAAH	0	0	0	0	0	0	1	UpstremRegulator
DOCK1	DOCK1	0	1	0	0	1	1	0	CommonTarget
ATP6V0A2	ATP6V0A2	1	0	0	0	0	0	0	EpigeneticEnzyme
MAFF	MAFF	0	0	0	0	0	0	0	PrimaryEpi
EID3	EID3	1	0	0	0	0	0	0	EpigeneticEnzyme
NAP1L3	NAP1L3	0	0	0	0	0	0	0	PrimaryEpi
MPO	MPO/DMR2	0	0	0	0	0	0	1	UpstremRegulator
EPCAM	EPCAM	0	0	1	1	0	0	0	EpigeneticEnzyme
ALDH2	ALDH2/DMR1	0	0	0	0	0	0	0	PrimaryEpi
ALDH2	ALDH2/DMR2	0	0	0	0	0	0	0	PrimaryEpi
HOXA9	HOXA9/DMR6	0	0	0	0	0	0	0	PrimaryEpi
SH3BP5	SH3BP5	1	0	0	0	0	0	0	EpigeneticEnzyme
SCRN1	SCRN1/DMR2	0	0	0	0	1	0	0	EpigeneticEnzyme
SCRN1	SCRN1/DMR3	1	0	0	0	0	0	0	EpigeneticEnzyme
HOXA5	HOXA5	0	0	0	0	0	1	0	UpstremRegulator
H1F0	H1F0	0	0	0	0	0	0	1	UpstremRegulator
RBPMS	RBPMS	0	0	0	0	0	0	0	PrimaryEpi
TRIP6	TRIP6	0	0	0	0	0	0	0	PrimaryEpi
HOXA7	HOXA7/DMR3	1	0	0	1	0	1	0	CommonTarget

OSBPL3	OSBPL3	0	0	0	0	0	0	0	PrimaryEpi
HOXA7	HOXA7/DMR4	0	0	0	0	0	1	0	UpstremRegulator
MOSC2	MOSC2	0	0	0	0	0	0	0	PrimaryEpi
IL12RB2	IL12RB2	0	0	0	0	0	0	0	PrimaryEpi
LAMB2	LAMB2	0	1	0	0	0	0	0	EpigeneticEnzyme
CYorf15A	CYorf15A	0	0	0	0	0	0	0	PrimaryEpi
PTGS2	PTGS2	0	0	0	0	0	0	0	PrimaryEpi
CRHBP	CRHBP	0	1	0	0	0	0	1	CommonTarget
LTB	LTB	1	0	0	0	0	0	0	EpigeneticEnzyme
SNCA	SNCA	0	0	0	0	0	0	0	PrimaryEpi
ERMP1	ERMP1	1	0	0	0	0	0	0	EpigeneticEnzyme
CHRDL1	CHRDL1	0	0	0	0	0	0	0	PrimaryEpi
GUCY1B3	GUCY1B3	0	0	0	0	0	0	0	PrimaryEpi
HOOK1	HOOK1	0	0	0	0	0	0	0	PrimaryEpi
LOC729046 /// RPL17	RPL17	0	0	0	0	0	0	0	PrimaryEpi
MPP6	MPP6	1	0	0	0	0	0	0	EpigeneticEnzyme
NFIA	NFIA	0	0	0	0	0	0	0	PrimaryEpi
SOCS2	SOCS2	0	0	0	0	0	0	1	UpstremRegulator
MREG	MREG	0	0	0	0	0	0	0	PrimaryEpi
CBR1	CBR1	0	0	0	0	0	0	0	PrimaryEpi
TMEM22	TMEM22	0	0	0	0	0	0	0	PrimaryEpi

C10orf10	C10orf10	0	0	0	0	0	0	0	PrimaryEpi
BGLAP	BGLAP	0	0	0	0	0	0	0	PrimaryEpi
MS4A3	MS4A3	0	1	1	1	0	0	1	CommonTarget
VNN1	VNN1	1	0	0	0	0	1	0	CommonTarget
LST1	LST1	1	0	0	0	0	1	0	CommonTarget
LPAR6	LPAR6	0	0	0	0	0	0	0	PrimaryEpi
FAM30A	FAM30A	0	0	0	0	0	0	1	UpstremRegulator
CD52	CD52	1	0	0	0	0	0	0	EpigeneticEnzyme
RXFP1	RXFP1	0	0	0	0	0	0	0	PrimaryEpi
BLNK	BLNK	0	0	0	0	0	1	0	UpstremRegulator
TNS3	TNS3	1	0	0	0	0	1	0	CommonTarget
CLIC2	CLIC2	0	0	0	0	1	1	0	CommonTarget
HDAC9	HDAC9	0	0	0	0	0	0	0	PrimaryEpi
		19	4	2	4	4	15	13	

Table 2.11. Multivariate analysis of overall survival of TCGA patients using either DNA methylation or gene expression

	DNA Methylation		Gene Expression	
Variable	HR(95% CI)	*p*	HR(95% CI)	*p*
Group	1.9(1.2-2.9)	0.003	1.7(1.0-2.7)	0.03
Age	1.04(1.03-1.06)	$8.5X\ 10^{-7}$	1.04(1.02-1.06)	$1\ X\ 10^{-6}$
Cytogenetic risk				
Intermediate vs low	2.7(1.3-5.2)	0.005	2.2(1.0-4.5)	0.04
High vs low	2.7(1.3-5.6)	0.006	2.2(1.0-4.7)	0.05
NPM1	0.8(0.5-1.3)	0.39	1.0(0.6-1.7)	1.00
FLT3	1.7(1.1-2.8)	0.03	1.5(0.9-2.5)	0.10

Table 2.12. Univariate overall survival analysis for LSC epigenetic signature regarding differential gene expression in various cohorts

	TCGA		**Metzeler et al**		**Wouters et al**		**Wilson et al**	
Variable	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*
LSC score (High vs. Low)*	2.4 (1.6-3.6)	$1x10^{-5}$	1.9 (1.3-2.8)	1×10^{-3}	2.3 (1.7-3.1)	$2x10^{-7}$	2.2 (1.6-3.1)	$2x10^{-6}$
Age	1.04 (1.03-1.06)	$1x10^{-9}$	1.03 (1.01-1.04)	3×10^{-4}	1.01 (1.0-1.03)	3×10^{-2}	1.03 (1.02-1.05)	$4x10^{-6}$
Cytogenetics								
Intermediate vs. low	2.7 (1.4-5.2)	2×10^{-3}	-	-	2.8 (1.7-4.7)	$3x10^{-5}$	1.9 (0.9-4.1)	1.1×10^{-1}
High vs. low	3.9 (1.9-7.8)	2×10^{-4}	-	-	4.7 (2.7-8.4)	$1x10^{-7}$	4.2 (1.9-9.6)	$6x\ 10^{-4}$
FLT3	1.1 (0.7-1.7)	7.1×10^{-1}	2.2 (1.5-3.3)	$8x10^{-5}$	1.8 (1.3-2.5)	5×10^{-4}	1.1 (0.8-1.6)	5.2×10^{-1}
NPM1	1.4 (0.9-2.1)	1.8×10^{-1}	0.8 (0.5-1.2)	2.4×10^{-1}	0.9 (0.6-1.3)	5.1×10^{-1}	0.8 (0.5-1.1)	2.0×10^{-1}

*LSC score is determined as described in method

Table 2.13. Multivariate overall survival analysis for LSC epigenetic signature regarding differential gene expression in various cohorts

	Metzeler et al		Wouters et al		Wilson et al	
Variable	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*
LSC score (High vs. Low)*	1.6 (1.0-2.4)	4×10^{-2}	1.8 (1.2-2.6)	3×10^{-3}	1.8 (1.2-2.6)	5×10^{-3}
Age	1.03 (1.01-1.04)	4×10^{-4}	1.02 (1.0-1.03)	3×10^{-2}	1.02 (1.01-1.04)	2×10^{-3}
Cytogenetics						
Intermediate vs. low	-	-	2.4 (1.4-4.2)	2×10^{-3}	1.4 (0.6-3.3)	3.9×10^{-1}
High vs. low	-	-	3.1 (1.6-5.9)	7×10^{-4}	3.0 (1.2-7.1)	2×10^{-2}
FLT3	2.3 (1.5-3.6)	1×10^{-4}	1.8 (1.3-2.6)	2×10^{-3}	1.6 (1.0-2.5)	4×10^{-2}
NPM1	0.7 (0.5-1.1)	8×10^{-2}	0.5 (0.3-0.7)	2×10^{-4}	0.7 (0.5-1.1)	1.2×10^{-1}

*LSC score is determined as described in method

Table 2.14. Univariate overall survival analysis for genetic mutations in epigenome modifying enzymes in TCGA

Genetic mutation	HR (95% CI)	*p*
DNMT3A	1.8 (1.2-2.7)	0.004
IDH1	0.8 (0.4-1.5)	0.4
IDH2	1.0 (0.6-1.9)	0.9
TET2	0.8 (0.4-1.7)	0.6
ASXL1	2.0 (0.6-6.4)	0.2

Table 2.15. Multivariate overall survival analysis including *DNMT3A* mutation for LSC epigenetic signature in TCGA

	DNA Methylation		Gene Expression	
Variable	HR(95% CI)	*p*	HR(95% CI)	*p*
Group	1.9 (1.2-3.0)	0.005	1.7 (1.0-2.7)	0.04
Age	1.0 (1.0-1.0)	9.3×10^{-7}	1.0 (1.0-1.0)	1.2×10^{-6}
Cytogenetic risk				
Intermediate/Normal	2.7 (1.3-5.6)	0.007	2.2 (1.0-4.6)	0.04
High	2.8 (1.3-5.9)	0.007	2.2 (1.0-4.8)	0.06
NPM1	0.8 (0.5-1.4)	0.46	1.0 (0.6-1.7)	0.99
FLT3	1.7 (1.0-2.8)	0.04	1.5 (0.9-2.5)	0.1
DNMT3	1.0 (0.6-1.6)	1.0	1.0 (0.7-1.6)	0.92

Table 2.16. Multivariate overall survival analysis for LSC epigenetic signature within intermediate cytogenetic risk patients in TCGA

	DNA Methylation		Gene Expression	
Variable	HR (95% CI)	*p*	HR (95% CI)	*p*
Group	1.8 (1.0-3.1)	0.05	1.9 (1.1-3.3)	0.03
Age	1.0 (1.0-1.1)	3×10^{-4}	1.0 (1.0-1.1)	2×10^{-4}
NPM1	0.7 (0.4-1.3)	0.2	0.9 (0.9-1.1)	0.7
FLT3	2.5 (1.3-4.8)	4×10^{-3}	2.3 (1.2-4.2)	0.01
DNMT3A	1.0 (0.6-1.8)	0.9	1.1 (0.7-1.9)	0.7

Table 2.17. Antibodies for Flow Cytometry

Cell Surface marker	Fluorophore	Manufacturer	Catalog Number	Working Dilution	Application
CD3	APC-Cy7	BD Bioscience	341090		To sort LSPCs from AML
CD19	PE-Cy5		555414	1:50	
CD20			555624		
CD34	APC		340667	1:50	
CD38	PE-Cy7		335790	1:100	
CD90	PE		555596	1:25	
CD3	APC-Cy7	BD Bioscience	341090	1:50	To test chimerism/ engraftment of LSC frequency in NSG mice
CD19	APC		555415	1:50	
CD33	PE		555450	1:50	
CD45	PB		560367	1:50	
CD45.1 (mouse)	PE-Cy7	eBioscience	25-0453-82	1:100	
Ter119 (mouse)	PE-Cy5		15-5921-83	1:100	

Table 2.18. Primers used for sequencing of *TET2*, *IDH1*, *IDH2,* and *DNMT3A* mutations of AML

Primers	Sequence 5' to 3'	Size	Tm	Reference
(1) TET2 exon 3 PCR1 F	TGAACTTCCCACATTAGCTGGT	955	55	Gelsi-Boyer *et al.* British Journal of Haematology 2009, 145(6): 788-800
(2) TET2 exon 3 PCR1 R	GAAACTGTAGCACCATTAGGCATT			
(3) TET2 exon 3 PCR1 Seq	GATAGAAATAAACACATTTT			
(4) TET2 exon 3 PCR2 F	CAAAAGGCTAATGGAGAAAGACGTA	836	55	
(5) TET2 exon 3 PCR2 R	GCAGAAAAGGAATCCTTAGTGAACA			
(6) TET2 exon 3 PCR3 F	GCCAGTAAACTAGCTGCAATGCTAA	846	55	
(7) TET2 exon 3 PCR3 R	TGCCTCATTACGTTTTAGATGGG			
(8) TET2 exon 3 PCR4 F	GACCAATGTCAGAACACCTCAA	867	60	
(9) TET2 exon 3 PCR4 R	TTGATTTTGAATACTGATTTTCACCA			
(10) TET2 exon 3 PCR5 F	TTGCAACATAAGCCTCATAAACAG	788	60	
(11) TET2 exon 3 PCR5 R	ATTGGCCTGTGCATCTGACTAT			
(12) TET2 exon 3 PCR6 F	GCAACTTGCTCAGCAAAGGTACT	781	60	
(13) TET2 exon 3 PCR6 R	TGCTGCCAGACTCAAGATTTAAAA			
(14) TET2 exon 4 F	ATACTACATATAATACATTCTAATTCCCTCACTG	495	55	
(15) TET2 exon 4 R	TGTTTACTGCTTTGTGTGTGAAGG			
(16) TET2 exon 5 F	CATTTCTCAGGATGTGGTCATAGAAT	286	55	
(17) TET2 exon 5 R	CCCAATTCTCAGGGTCAGATTTA			
(18) TET2 exon 6 F	AGACTTATGTATCTTTCATCTAGCTCTGG	599	60	
(19) TET2 exon 6 R	ACTCTCTTCCTTTCAACCAAAGATT			
(20) TET2 exon 7 F	ATGCCACAGCTTAATACAGAGTTAGAT	362	55	
(21) TET2 exon 7 R	TGTCATATTGTTCACTTCATCTAAGCTAAT			

(22) TET2 exon 8 F	GATGCTTTATTTAGTAATAAAGGCACCA	354	55	
(23) TET2 exon 8 R	TTCAACAATTAAGAGGAAAAGTTAGAATAATATTT			
(24) TET2 exon 9 F	TGTCATTCCATTTTGTTTCTGGATA	361	55	
(25) TET2 exon 9 R	AAATTACCCAGTCTTGCATATGTCTT			
(26) TET2 exon 10 F	CTGGATCAACTAGGCCACCAAC	774	55	
(27) TET2 exon 10 R	CCAAAATTAACAATGTTCATTTTACAATAAGAG			
(28) TET2 exon 11 PCR1 F	GCTCTTATCTTTGCTTAATGGGTGT	748	60	
(29) TET2 exon 11 PCR1 R	TGTACATTTGGTCTAATGGTACAACTG			
(30) TET2 exon 11 PCR2 F	AATGGAAACCTATCAGTGGACAAC	1107	60	
(31) TET2 exon 11 PCR2 R	TATATATCTGTTGTAAGGCCCTGTGA			
(32) IDH1 exon 4 F	TGTGTTGAGATGGACGCCTATTTG	481	55	Thol F *et al.* Haematologica 2010, 95(10): 1668-1674
(33) IDH1 exon 4 R	TGCCACCAACGACCAAGTCA			
(34) IDH2 exon 4 F	GGGGTTCAAATTCTGGTTGA	290	53	
(35) IDH2 exon 4 R	CTAGGCGAGGAGCTCCAGT			
(36) DNMT3A exons 7-8 F	ATGGTCCCCTTGAGTGTCAG	836	56	Fernandez-Mercado *et al.* PLoS One 2012, 7(8): e42334
(37) DNMT3A exons 7-8 R	CATCACCCCAATTCCAGACT			
(38) DNMT3A exons 9-10 F	CTGTATCTGGTCCCCTCCAG	747	56	
(39) DNMT3A exons 9-10 R	CTCCCTAAGCATGGCTTTCC			
(40) DNMT3A exons 11-12 F	GGGAACAAGTTGGAGACCAG	490	56	
(41) DNMT3A exons 11-12 R	GGTCCCATGTCATTCAAACC			
(42) DNMT3A exon 13 F	GTCACAGTGCCTCCCTTTTC	308	56	
(43) DNMT3A exon 13 R	TGGACACAGTCAGCCAGAAG			
(44) DNMT3A exon 14 F	CAGGGCTTAGGCTCTGTGAG	359	56	
(45) DNMT3A exon 14 R	AGGTGTGCTACCTGGAATGG			
(46) DNMT3A exons 15-16 F	CGGTCTTTCCATTCCAGGTA	614	56	

(47) DNMT3A exons 15-16 R	CATCATTTCGTTTTGCCAGA			
(48) DNMT3A exon 17 F	GACTTGGGCCTACAGCTGAC	345	58	
(49) DNMT3A exon 17 R	CAAAATGAAAGGAGGCAAGG			
(50) DNMT3A exons 18-19 F	CTTCCTGTCTGCCTCTGTCC	552	56	
(51) DNMT3A exons 18-19 R	ATGAAGCAGCAGTCCAAGGT			
(52) DNMT3A exons 19b-20 F	GCAGCACTGTGCAATATGGT	549	56	
(53) DNMT3A exons 19b-20 R	CTTCCCCACTATGGGTCATC			
(54) DNMT3A exons 21 F	GCGGGGAGTTTGAAGAGAGT	342	56	
(55) DNMT3A exons 21 R	CCACACTAGCTGGAGAAGCA			
(56) DNMT3A exons 22 F	TTTGGTAGACGCATGACCAG	301	56	
(57) DNMT3A exons 22 R	CAGGACGTTTGTGGAAAACA			
(58) DNMT3A exons 23 F	TCCTGCTGTGTGGTTAGACG	654	56	
(59) DNMT3A exons 23 R	CCTCTCTCCCACCTTTCCTC			
(60) DNMT3A exon 17 F	CCTCGATGTCCTTACTATGGATACTCCA	402	63	Additional primers designed to cover the ones not working in previous rows, all the three new pairs worked on DNMT3A exon 17
(61) DNMT3A exon 17 R	CAAGGGCTGCCTCCAGGTGCTGAG		69	
(62) DNMT3A exon 17 F	CTCACCTGCCGAGACCAG	276	59	
(63) DNMT3A exon 17 R	CCTCCAGGTGCTGAGTGTG		60	
(48) DNMT3A exon 17 F	GACTTGGGCCTACAGCTGAC	437	60	
(64) DNMT3A exon 17 R	TTTGCCCTTTACCCTCTCAA		57	

Note: For IDH1 and IDH2, a single point mutation was tested in exon 4 (R132 and R140 respectively); for TET2 and DNMT3A mutations, multiple exons were tested based on regions of frequent somatic mutation according to COSMIC database (Wellcome Trust Sanger Institute).

Chapter 3

Epigenetic basis of human normal hematopoietic development

Summary

DNA methylation plays an indispensable role during tissue development and cellular differentiation. Hematopoiesis is one of best-understood and characterized developmental processes. We hypothesized that DNA methylation would be an essential mechanism during lineage differentiation in human hematopoiesis. Here, we provide a comprehensive methylome map of HSPCs, particularly in myeloid lineage, by performing a genome-wide DNA methylation analysis. We found that DNA methylation distinguished distinct lineages of hematopoiesis, suggesting a critical role of it during hematopoietic development. In concordance with previous studies for tissue development, we observed that most of DNA methylation changes during hematopoietic differentiation occurred at non-CpG island regions such as CpG island shores or open seas. We identified DMRs in potential novel regulators of hematopoiesis including *HMHB1* and *MIR539,* as well as in previously known genes such as *MPO* and *CDK6.* The DMRs for normal hematopoiesis were enriched in a regulatory genomic element, super-enhancer. We found massive epigenetic variation between murine and human hematopoiesis, explaining the intrinsic differences during the hematopoietic development in those species.

Results

Comprehensive DNA Methylation Analysis Shows Tight Clustering of Human Hematopoietic Stem and Progenitor Cells

Human hematopoiesis proceeds through a series of multipotent and oligopotent stem and progenitor cells that progressively lose self-renewal ability and become more restricted in their differentiation potential (Figure 3.1). These critical functional properties are mediated in part through epigenetic mechanisms, including DNA methylation. We obtained bone marrow from five normal donors and isolated HSPC by fluorescence-activated cell sorting (FACS) including: hematopoietic stem cells (HSC), multipotent progenitors (MPP), L-MPP, common myeloid progenitors (CMP), megakaryocyte/erythroid progenitors (MEP), and GMP (Figure 3.2, Tables 3.1 and 3.2).

In order to further understand epigenetic variation during early human hematopoiesis, we generated genome-scale methylation profiles for normal hematopoietic stem and progenitor cell populations. Strikingly, multidimensional scaling analysis utilizing the top 1000 most variable CpG positions revealed tight clustering of human HSPC populations by lineage with no outliers (Figure 3.3). As the distance between clusters in multidimensional scaling is a measure of their similarity, this analysis indicates that DNA methylation reflects the function and hierarchy of the HSPC populations. For example, HSC and MPP clusters are close together reflecting their functional similarities. Similarly, L-MPP are located between HSC/MPP and GMP clusters, but farther away from MEP, supporting the hypothesis that L-MPP is an early lymphoid progenitor, which retains myeloid programs for GMP, but not MEP, differentiation (Doulatov et al., 2010; Goardon et al., 2011).

DMR Analysis Identified Previously Known and Novel Regulators in Human Hematopoietic Development

The DMRs identified across HSPC not only permit clustering of these populations, but also have the potential to reveal novel regulators of hematopoietic lineage development. We first examined the DMRs for genes already known to play such a role. The DMR analysis identified myeloperoxidase (*MPO)*, a well-established protein involved in neutrophil activity (Klebanoff, 2005), which showed progressive hypomethylation going from HSC to GMP, but exhibited hypermethylation in MEP (Figure 3.4). Similarly, *CDK6*, a cyclin-dependent protein kinase important in hematopoietic cell differentiation (Kozar and Sicinski, 2005; Malumbres et al., 2004), was progressively hypermethylated during differentiation from HSC to MEP and GMP (Figure 3.4). In both cases, these results were confirmed through direct pyrosequencing of these loci (Figure 3.4), validating the DNA methylation array approach. In addition to these two examples, a number of genes known to be involved in early hematopoiesis were found to be differentially methylated including: *STAT3, KEL, IDH2, HLF, NOTCH1*, *GATA1*, and *HOX* family genes (Table 3.3, see Appendix2).

This analysis also identified novel sites of epigenetic variation during hematopoiesis. For example, *HMHB1*, encoding one of the minor histocompatibility antigens, was found to be hypomethylated in L-MPP and GMP, suggesting a possible role in GMP differentiation (Figure 3.5). Progressive hypomethylation was also identified in *MIR539* going from HSC to MEP, suggesting that this microRNA may contribute to erythropoiesis (Figure 3.5). Interestingly, the *MIR539* gene is located in DLK1-DIO3 imprinting region that contains a miRNA cluster involved in leukemia pathogenesis

(Benetatos et al., 2013). Further validation of these novel candidate regulators will require functional experiments.

We next sought to classify the DMRs among normal HSPCs according to their global genomic location in islands, shores, shelves, and open seas. Focusing on the comparison of HSC with GMP, we found that most DMRs were not in CpG islands, but were enriched outside of the islands, predominantly in shelves and open seas, compared to the distribution of random length matched DMRs (Figure 3.6a). The comparison of HSC with MEP also showed enrichment of DMRs at non-CpG island regions such as shores, shelves, and open seas (Figure 3.6b). Similarly, genes with an inverse correlation between DNA methylation and gene expression were located outside of the islands themselves, with the strongest correlation at shores and open seas for both comparisons (Figure 3.7). In addition to these comparisons, more than 50% of the DMRs among HSPCs were in open seas (Table 3.4). Thus, functional epigenetic differences during early human hematopoietic differentiation occur in CpG sparse regions, consistent with other recent studies of differentiation (Irizarry et al., 2009; Ji et al., 2010) and cancer (Hansen et al., 2011; Irizarry et al., 2009).

DMRs for Normal Hematopoiesis are Enriched in Super-enhancers

In order to further investigate a direct mechanistic link between DNA methylation change and regulation of hematopoietic development, we examined the overlap of DMRs that we identified for normal hematopoiesis in DNA regulatory elements, super-enhancers. Super-enhancers are clusters of transcriptional enhancers bound by mater transcription factors, and associated with genes that define cell identity and tissue types (Hnisz et al., 2013; Whyte et al., 2013). From a published study that has identified super-

enhancer in 86 different human cell and tissue samples, we selected three categories of tissue types; first, normal cell types related to hematopoietic development such as CD34+ (hematopoietic progenitor cells) and CD14+ (monocytes), second, cell lines related to hematological disorders including jurkat, K562 and MM1S, third, other normal tissues or cell lines including H1, adipose, angular gyrus, and spleen. First, we examined the enrichment of DMRs from normal hematopoiesis including HSC vs GMP and HSC vs MEP comparisons, in super-enhancers of the different tissues or cell types. Remarkably, DMRs for GMP differentiation were significantly enriched in relevant cell types including monocytes and CD34+ cell population (Table 3.5). We identified master transcription factors showing differential methylation as HSC differentiates into GMP, such as *ZNF217*, known to be involved in cell proliferation (Figure 3.8a and Table 3.6, see Appendix3). Among the DMRs of HSC vs GMP comparison, contained within super-enhancers of monocytes, 27.9% was located in monocyte (CD14+ cells) specific super-enhancers (Table 3.6, see Appendix3). For example, a DMR of *TREM1* that encodes a receptor expressed on monocytes was located in a monocyte specific super-enhancer and showed hypomethylation along with upregulation in GMP (Figure 3.8b and Table 3.6, see Appendix3). In addition to the relevant cell types, the DMRs of HSC vs GMP were overlapped with super-enhancers of jurkat and MM1S cells whereas, DMRs distinguishing MEP from HSC are enriched in K562 cells, implying involvement of epigenetic deregulation of the regulatory region in disease pathogenesis.

Human Hematopoiesis Displays Distinct Epigenetic Regulation Compared to Murine Hematopoiesis

In addition to these genomic regional changes, there was an overall global change in the level of DNA methylation in human hematopoiesis (Figure 3.9 and Table 3.4). Global hypomethylation was observed upon MEP differentiation from CMP, but not in GMP differentiation, and most DMRs distinguishing MPP from CMP were less methylated in CMP in human hematopoiesis (Figure 3.9 and Table 3.4). We also found that all DMRs lost methylation when L-MPP differentiates into GMP (Figure 3.9 and Table 3.4). Previous epigenetic study of mouse hematopoiesis (Ji et al., 2010) allowed us to compare human and mouse hematopoietic differentiation for two pairs, MPP vs CMP and CMP vs GMP. We examined overlap of genes of DMR lists of the two comparisons. In both comparisons, most of the genes were different with only 4-5% of total genes in human overlapping with mouse: 7 out of 166 genes of MPP vs CMP, and 30 out of 728 genes of CMP vs GMP (Table 3.7, see Appendix4). Among the 7 and 30 common genes for the two comparisons, 5 and 14 of them showed the opposite direction of methylation changes for the MPP vs CMP and CMP vs GMP pairs, respectively. For example, *VKORC1L1* that encodes a subunit of the vitamin K epoxide reductase complex, was hypermethylated when CMP differentiate into GMP in human, but hypomethylated in the CMP to GMP transition in mouse (Table 3.7, see Appendix4). These results suggest that there is great variability in the specific epigenetic sites regulating mouse and human hematopoiesis.

Discussion

The role of DNA methylation in hematopoietic development has been implicated in several studies that have shown an indispensable role of DNMT1, DNMT3A, and

DNMT3B in hematopoiesis (Broske et al., 2009; Challen et al., 2012; Tadokoro et al., 2007; Trowbridge et al., 2009). In addition, dynamic DNA methylation changes during hematopoietic development in mouse and human have been demonstrated (Bartholdy et al., 2014; Bock et al., 2012; Ji et al., 2010). In order to provide a comprehensive methylome map of human hematopoiesis, particularly myeloid lineage, we analyzed genome-wide DNA methylation of a complete set of HSPCs in myeloid lineage. Our study suggested several significant findings for the epigenetic basis of human hematopoietic development. First, global hypomethylation is a core mechanism of hematopoietic differentiation, as MPP to CMP, CMP to MEP, and L-MPP to GMP differentiations were accompanied by loss of methylation (Figure 3.9). In addition to identifying previously known and novel candidate genes during hematopoietic differentiation, we found that DNA methylation were likely to occur at regulatory elements such as super-enhancer, suggesting a critical role of epigenetics in gene expression of lineage specific genes. Furthermore, we found distinct epigenetic plasticity between human and mouse hematopoiesis. We observed few overlaps of genes with DNA methylation changes between mouse and human hematopoiesis. Besides the distinct study design and experimental platform between the studies, intrinsically distinct mechanisms exist in hematopoietic development in mouse and human. For example, human blood contains 3 to 5-fold more neutrophils than mouse, while mouse blood has 2 fold more lymphocytes, indicating differences in the generation of specific types of hematopoietic progenitor cells (Mestas and Hughes, 2004). Furthermore, mouse hematopoietic system is composed of a different set of HSPCs, and a different set of cell surface markers for HSPCs are used to isolate each population. For example, multiple

subpopulations constitute an MPP compartment, and Flk2 and Slamf1 are used to differentiate those subpopulations (Bock et al., 2012; Chao et al., 2008; Doulatov et al., 2012; Ji et al., 2010).

We could compare our results to previous DNA methylation study of human hematopoiesis, and found few common observations, yet there were many distinct and novel findings in this study. We observed global hypomethylation when MPPs differentiated into CMPs, as Bartholdy et al. has reported. However, we identified global hypomethylation in CMP to MEP transition, while Bartholdy et al. has observed balanced hyper and hypomethylation in this differentiation (Bartholdy et al., 2014). In addition, we noticed that DMR lists from all pair-wise comparisons among HSPCs ($p < 0.01$) of our study significantly differed from 561 loci distinguishing LT-HSC, ST-HSC, CMP, and MEP, identified in Bartholdy et al. 51 out of 3613 (1.4%) DMRs of our lists including DMRs in *KLF1* and *GATA1*, overlapped with the 561 loci from Bartholdy et al. The discordance between our study and Bartholdy et al. may come from two major sources: a different set of HSPCs and methods used in those studies. Our study includes two more progenitor populations, L-MPP and GMP, beside HSC, MPP (called as ST-HSC in Bartholdy et al.), CMP, and MEP. We have used 450K array which interrogate ~48000 CpGs across the genome without a bias toward CpG islands and gene promoters, while HELP_Promoter array, used in Bartholdy et al., examines ~26000 loci, primarily targeting gene promoters.

In summary, we provide a comprehensive methylome map of a complete set of HSPCs in myeloid lineage, demonstrating the role of DNA methylation at regulatory

elements during lineage differentiation and dynamic epigenetic plasticity distinguishing human hematopoiesis from murine hematopoietic development.

Materials and Methods

Human Samples

Fresh human bone marrow mononuclear cells (BMMC) from healthy donors (2 x 10^8 cells per donor, Catalog#: ABM006) were purchased from ALLCELLS® (Emeryville, CA). A CD34+ cell-enrichment step was performed with the human progenitor cells enrichment kit with CD61 depletion (Stem Cell Technologies, Canada, Catalog # 19356) on a RobSep machine from the same company. PBMC or BMMC were separated with Ficoll-Paque Plus (Amersham Biosciences, Piscataway, NJ, Catalog number: 17-1440-03), and cryopreserved in 1 x freezing medium (90%FBS + 10%DMSO).

Flow Cytometry Analysis and Cell Sorting

A battery of antibodies (Abs) was used for staining, analysis and sorting of progenitor cells from either healthy BMMCs, as well as lineage analysis human chimerism/engraftment (Table 3.2). Cells were either analyzed or sorted using a FACS Aria II cytometer (BD Biosciences, Franklin Lakes, NJ). Analysis of flow cytometry raw data was done with FlowJo Software (Treestar, Ashland, OR).

Illumina Infinium Human Methylation 450 Bead Array Assay

Genomic DNA from each sample was purified using the MasterPure DNA purification kit (Epicentre) according to the manufacturer's protocol. The genomic DNA (250-500ng) was treated with sodium bisulfate using the Zymo EZ DNA Methylation Kit (ZYMO Research) as recommended by the manufacturer, with the alternative incubation

conditions for the Illumina Infinium Methylation Assay. Converted DNA was eluted in 11ul of elution buffer. DNA methylation level was measured using Illumina Infinium HD Methylation Assay (Illumina) according to the manufacturer's specifications. Methylation array data are deposited at the Gene Expression Omnibus (GEO) with accession number GSE63409.

Illumina Infinium Human Methylation 450 Bead Array Analysis

Raw intensity files were obtained using minfi package (Aryee et al., 2014) to calculate methylation ratios (Beta values). The data was normalized using Illumina preprocessing method implemented in minfi. Several quality control measures were applied to remove arrays with low quality. Control probes were examined on the 450k array to assess several measures including bisulfite conversion, extension, hybridization, specificity and others. One of the MPP samples (BM2712) showed low quality for the measures, so they were removed for further analysis. Next, median methylated and unmethylated signals were calculated for each arrays; no array was identified for signal values lower than 10.5. For multidimensional scaling analysis, probes containing an annotated SNP (dbSNP137) at the single-base extension or CpG sites were removed (17398 probes removed). Minfi 1.8.9 was used.

Bump hunting method previously described was applied to identify DMRs in 450k array (Aryee et al., 2014; Jaffe et al., 2012). Beta value of 0.1 (10% of methylation difference) was used as cutoff when finding DMRs. Statistical significance was assigned by permutations testing and the P-value cutoff used for downstream analysis was <0.01 that corresponded to Benjamini-Hochberg adjusted p-value <0.1 (data not shown) unless different cutoff was designated in result part. Bumphunter 1.2.0 was used.

Bisulfite Pyrosequencing

100ng of genomic DNA from each sample was treated with sodium bisulfate using an EZ DNA methylation Gold Kit (ZYMO research) following manufacturer's protocol. The bisulfate treated DNA was PCR amplified using unbiased nested primers. Quantitative pyrosequencing was performed using a PSQ HS96 (Biotage) to validate DMR regions. The DNA methylation percentage at each CpG site was measured using the Q-CpG methylation software (Biotage). SssI treated human genomic DNA was used as 100% methylated controls and human genomic DNA amplified by Repli-G mini kit (Qiagen) was used as the non-methylated (0%) DNA control. Table 3.8 provides the primer sequence used for the pyrosequencing reactions with the chromosomal coordinates in the University of California at Santa Cruz February 2009 human genome assembly (hg19) for each CpG site investigated.

Affymetrix Microarray Expression Analysis

Total RNA was extracted from each FACS-sorted cell population using RNeasy® Plus Mini (QIAGEN, Valencia, CA, Catalog#: 74134) according to the manufacture's protocol. All RNA samples were quantified with 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), subjected to reverse transcription, two consecutive rounds of linear amplification, and production and fragmentation of biotinylated cRNA. 15μg of cRNA from each sample was hybridized to HG U133 Plus 2.0 microarrays. Hybridization and scanning were performed according to the manufacture's instruction (Affymetrix). This step was performed at the PAN center of Stanford University. Data were normalized by GC robust multi-array average method and analyzed on R/Bioconductor. BM2770 GMP, BM 2759 L-MPP, BM2761 CMP, BM2770 CMP were

removed from further analysis due to low quality (GEO GSE63270). University of California at Santa Cruz February 2009 human genome assembly (hg19) for each CpG site investigated.

Affymetrix Microarray Expression Analysis

Total RNA was extracted from each FACS-sorted cell population using RNeasy® Plus Mini (QIAGEN, Valencia, CA, Catalog#: 74134) according to the manufacture's protocol. All RNA samples were quantified with 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), subjected to reverse transcription, two consecutive rounds of linear amplification, and production and fragmentation of biotinylated cRNA. 15μg of cRNA from each sample was hybridized to HG U133 Plus 2.0 microarrays. Hybridization and scanning were performed according to the manufacture's instruction (Affymetrix). This step was performed at the PAN center of Stanford University. Data were normalized by GC robust multi-array average method and analyzed on R/Bioconductor. BM2770 GMP, BM 2759 L-MPP, BM2761 CMP, BM2770 CMP were removed from further analysis due to low quality (GEO GSE6327).

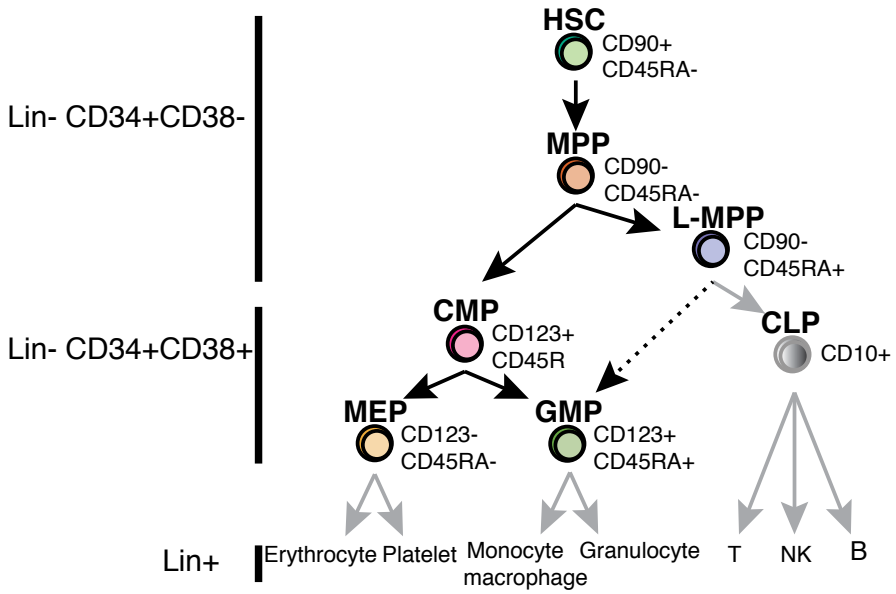


Figure 3.1. Schematic of human hematopoiesis with the immunophenotype of individual HSPC populations as indicated. Note the color scheme for each HSPC population is used throughout.

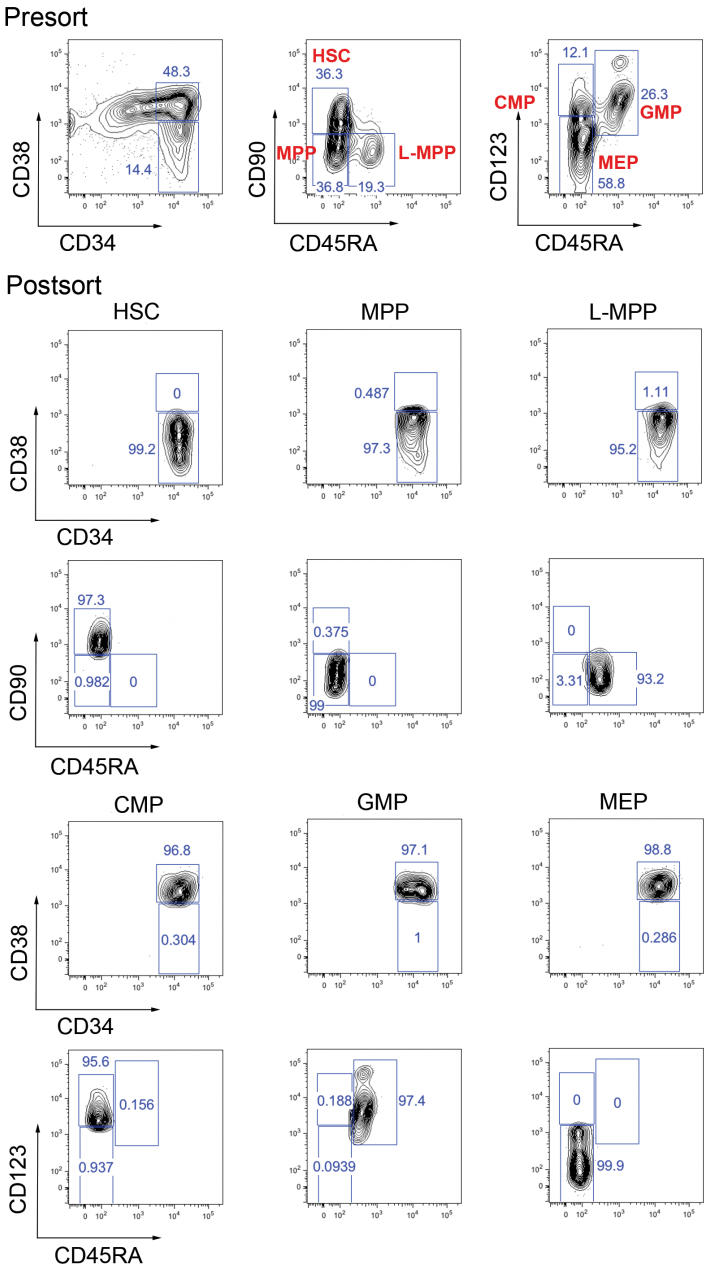


Figure 3.2. Pre-sort and post-sort FACS analysis of HSPCs from human bone marrow. Top panel: FACS-sorting scheme of six populations of HSPCs from normal human BM. Other panels: The second round of post-sort analysis to check the purity of sorting.

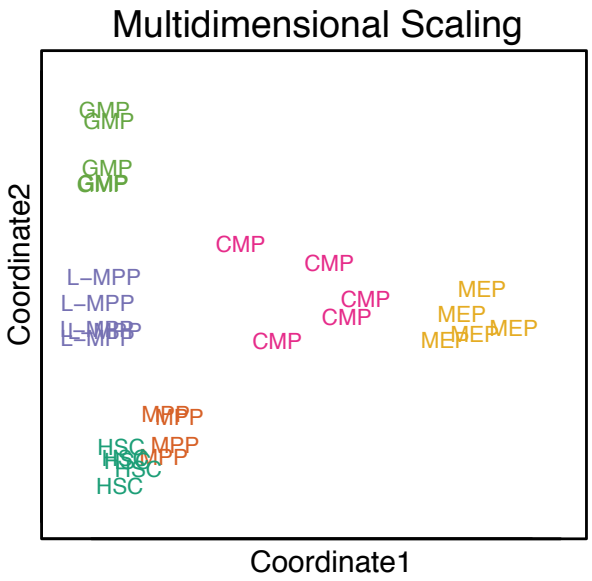


Figure 3.3. Comprehensive DNA methylation analysis shows tight clustering of HSPCs by their lineages. Multidimensional scaling examining the top 1,000 most variable methylation positions among normal progenitors shows tight clustering of distinct lineages.

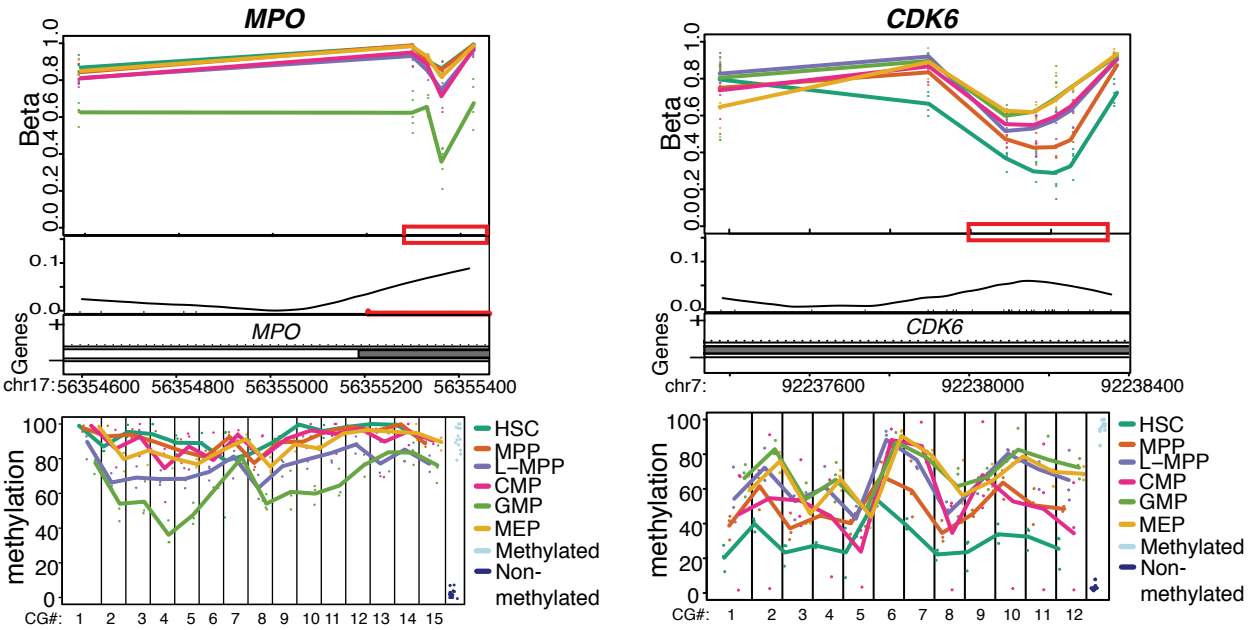


Figure 3.4. DMR plots indicating genomic loci for genes with previously known functions in hematopoiesis *MPO* and *CDK6*. Top: level of CpG methylation (beta) of each sample for the region; Middle: CpG density (curve), CpG sites (black tick marks), CpG islands (red lines); Bottom: gene annotation; Lower panel: bisulfite pyrosequencing replicating the methylation value for individual CpGs in the red boxes.

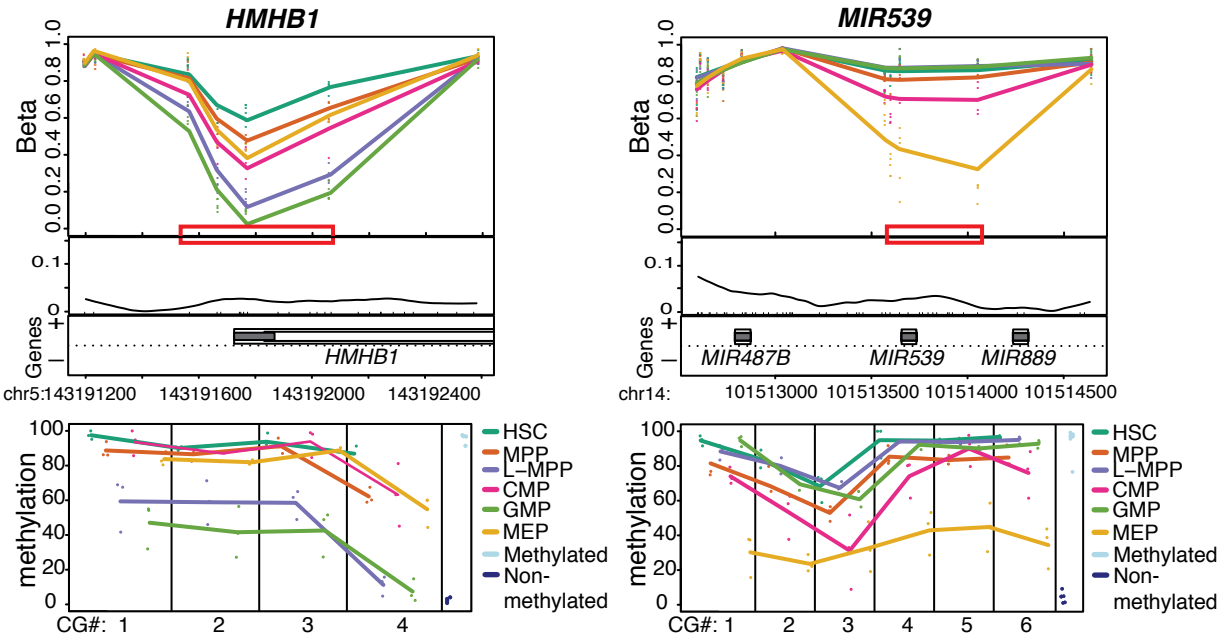


Figure 3.5. DMR plots indicating genomic loci for newly identified genes with previously unknown functions in hematopoiesis *HMHB1* and *MIR539*. Top: level of CpG methylation (beta) of each sample for the region; Middle: CpG density (curve), CpG sites (black tick marks), CpG islands (red lines); Bottom: gene annotation; Lower panel: bisulfite pyrosequencing replicating the methylation value for individual CpGs in the red boxes.

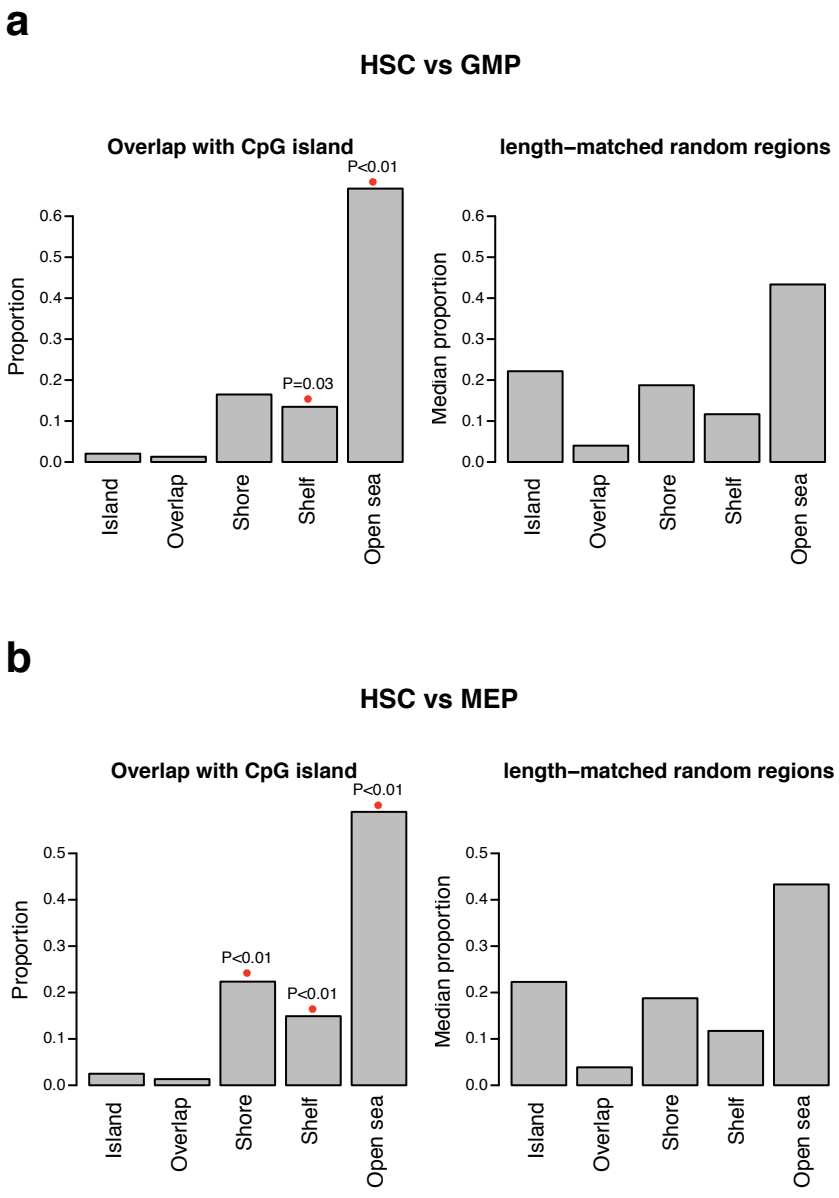


Figure 3.6. Location of DMRs for normal hematopoiesis relative to CpG island. Left panel shows proportion of DMRs located in CpG island, overlapping region (50% overlap with island), shore, shelf, and open sea. Right panel shows distribution of length-matched random regions relative to CpG island, overlap, shore, shelf, and open sea. (**a**) HSC vs GMP. (**b**) HSC vs MEP.

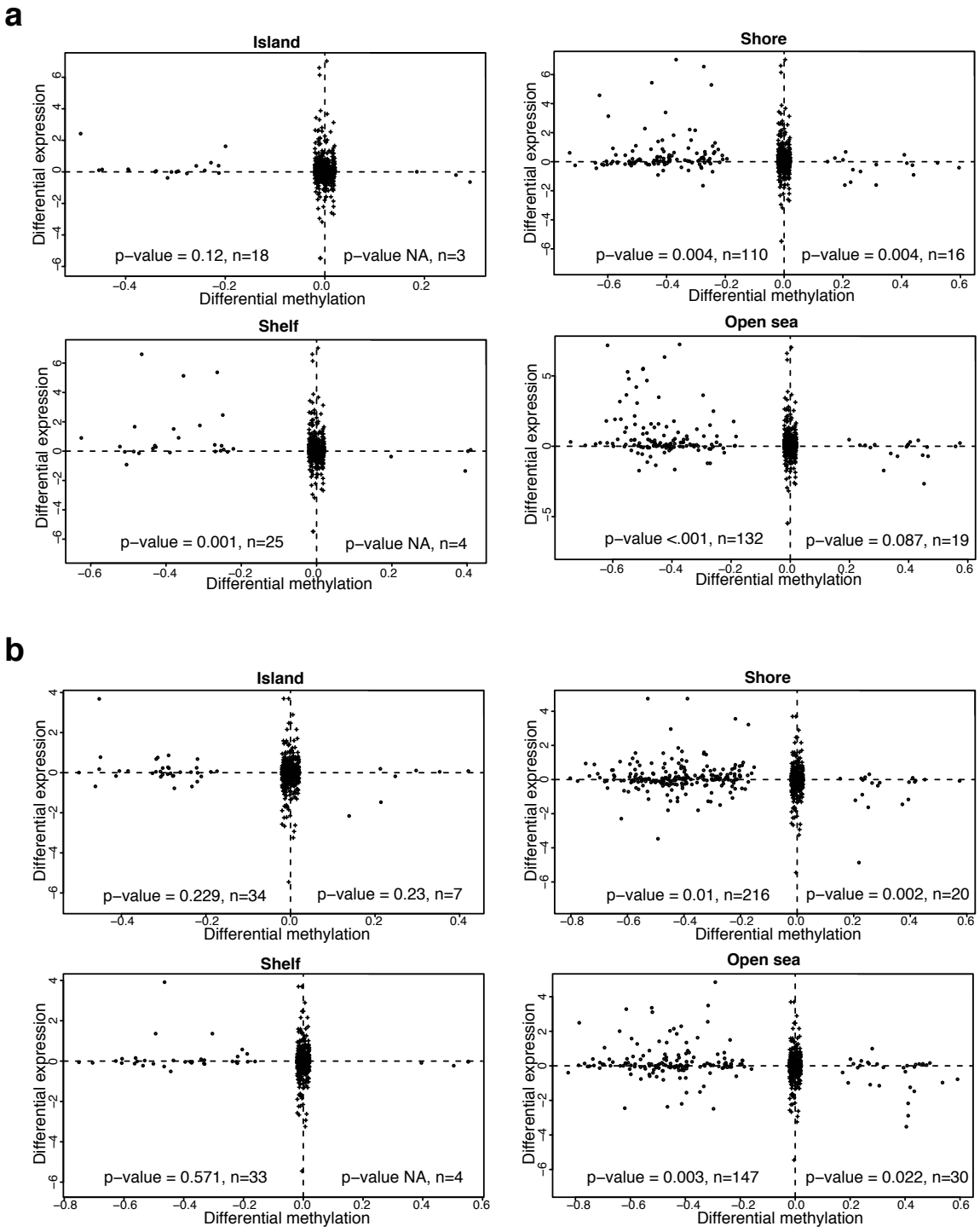

a
Island
Differential expression
Differential methylation
p−value = 0.12, n=18
p−value NA, n=3
Shore
p−value = 0.004, n=110
p−value = 0.004, n=16
Shelf
p−value = 0.001, n=25
p−value NA, n=4
Open sea
p−value <.001, n=132
p−value = 0.087, n=19
b
Island
p−value = 0.229, n=34
p−value = 0.23, n=7
Shore
p−value = 0.01, n=216
p−value = 0.002, n=20
Shelf
p−value = 0.571, n=33
p−value NA, n=4
Open sea
p−value = 0.003, n=147
p−value = 0.022, n=30

Figure 3.7. Gene expression inversely correlates with DMRs at non-CpG island regions in normal hematopoiesis. DMRs located within 2kb of gene TSSs (black dots) were classified into 4 groups according to the distance relative to CpG island: island, shore, shelf, and open sea. DMRs located further than 2kb of gene TSSs are denoted as black pluses in the middle. Log2 ratios of differential expression were plotted against differential methylation (all values are from group2-group1). Wilcoxon rank-sum test was performed to test the null hypothesis that the expression differences for the hypo- or hypermethylated DMRs within 2kb of gene TSSs (black dots) showed stronger inverse correlation than the expression differences of the random DMRs that are located further than 2kb of TSSs (black pluses). Random DMRs were shown in the middles of DNA methylation axis regardless of their methylation differences. (**a**) For HSC vs GMP, shore showed statistically inverse correlation of DMR with gene expression. (**b**) For HSC vs MEP, shore and open sea showed statistically inverse correlation of DMR with gene expression.

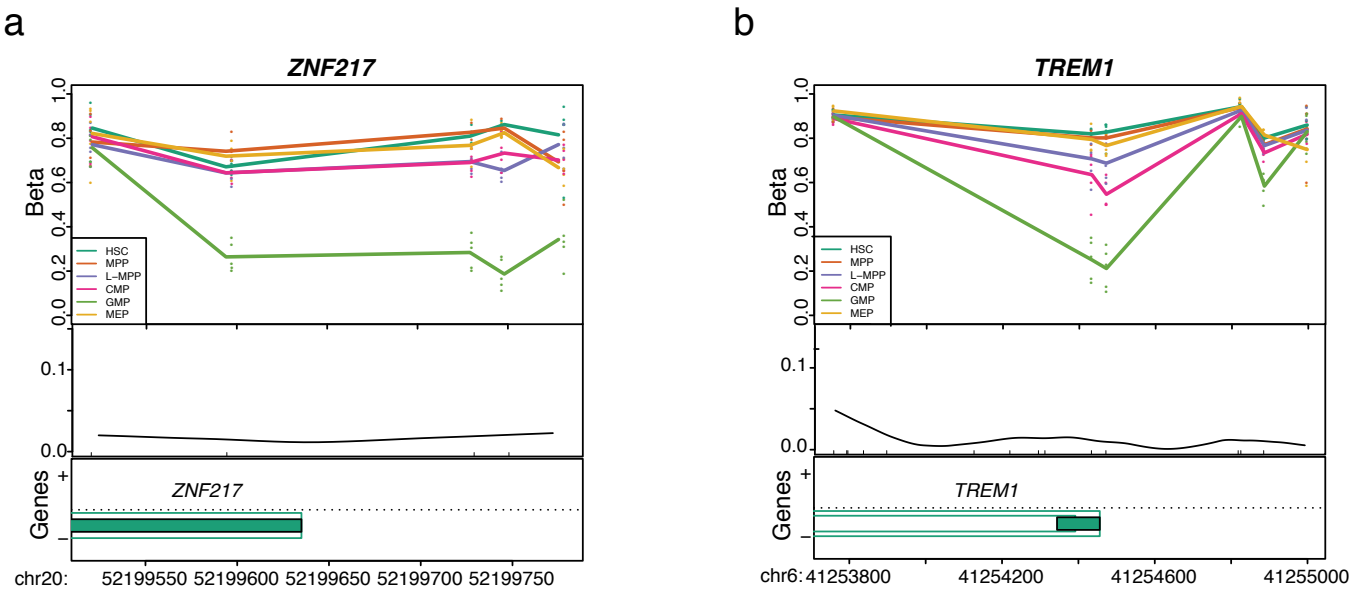


Figure 3.8. Examples of DMRs located in master transcription factor or super-enhancer. (**a**) *ZNF217*, a master transcription factor in monocyte differentiation. (**b**) *TREM1*, located in monocyte specific super-enhancer.

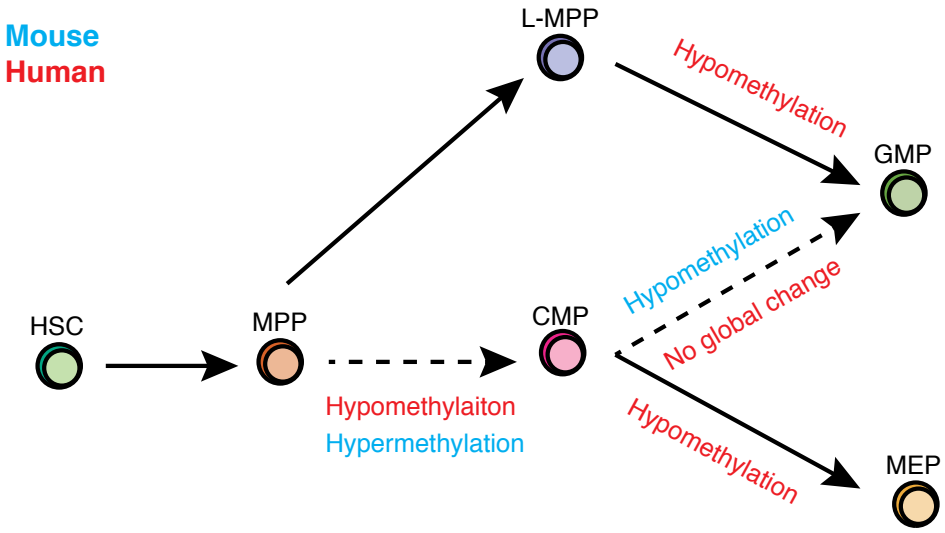


Figure 3.9. Global methylation changes during hematopoietic development in human and mouse. The direction of global methylation change is shown for each comparison in red for human and blue for mouse hematopoiesis. The dotted lines represent two comparisons shared between the studies for human and mouse hematopoiesis.

Table 3.1. Normal bone marrow donor sample analysis

Sample ID	Age	Gender	Application
BM2627	30	M	450K
BM2710	29	F	450K
BM2712	39	M	450K
BM2748	24	M	450K
BM2753	22	F	450K
BM2759	38	M	GEP
BM2761	26	M	GEP
BM2768	25	M	GEP
BM2770	39	F	GEP
BM2793	18	F	GEP
BM2794	21	M	GEP
BM2806	35	M	GEP
BM3604	26	M	P
BM3668	24	M	P
BM3671	24	M	P

Abbreviations: 450K, Illumina Infinium Human Methylation 450K BeadChip array; F, female; GEP, gene expression profiling microarray; M, male; P, bisulfite pyrosequencing

Table 3.2. Antibodies for flow cytometry

Cell Surface marker	Fluorophore	Manufacturer	Catalog Number	Working Dilution	Application
CD2	PE-Cy5	BD Bioscience	555328	1:50	To sort normal HSPCs from BMs
CD3			555341		
CD4			555348		
CD7			555362		
CD8			555368		
CD10			555376		
CD11b			555389		
CD14			340585		
CD19			555414		
CD20			555624		
CD56			555517		
CD235a			559944		
CD34	APC		340667	1:50	
CD38	PE-Cy7		335790	1:100	
CD45RA	PB		560362	1:25	
CD90	FITC		555595	1:25	
CD123	PE		554529	1:25	

Table 3.3. DMR lists for pair-wise comparisons among HSPCs (**See Appendix2**)

Table 3.4. Summary of DMRs identified in the indicated pairwise comparisons

Comparisons (Group1 versus Group2)	Numbers of DMRs*		Locations of DMRs relative to CpG islands (%)			
	Group1>Group2	Group1<Group2	Islands	Shores	Shelves	Open seas
HSC vs MPP	14	2	33.3	0	0	66.7
MPP vs CMP	158	8	3.6	22.3	10.2	63.9
CMP vs GMP	366	362	2.5	17.0	15.4	65.1
CMP vs MEP	319	13	3.6	25.6	16.0	54.8
GMP vs MEP	1308	764	2.4	19.3	15.3	63
MPP vs L-MPP	49	54	2.9	15.5	19.4	62.1
HSC vs L-MPP	165	109	2.2	19.0	15.0	63.7
L-MPP vs GMP	556	0	2.2	15.5	13.3	69.1
HSC vs GMP	1168	162	2.0	17.8	13.5	66.7
HSC vs MEP	1545	190	2.5	23.7	14.9	58.9

Table 3.5. Enrichment of DMRs for normal hematopoiesis in super-enhancers

Tissues or Cell lines	Cell type	DMR lists*	
		HSC vs GMP	HSC vs MEP
CD34+(adult)	HSPC	0.009	2.5×10^{-9}
CD14	Monocyte	6.9×10^{-16}	0.12
Jurkat	T cell leukemia	5.3×10^{-6}	1
K562	Erythroleukemia(CML)	1	9.6×10^{-9}
MM1S	Multiple myeloma	0.04	0.08
Adipose	Adipose	1	1
Angular gyrus	Brain	0.15	0.66
H1	Embryonic stem cell	0.90	0.78
Spleen	Spleen	1.1×10^{-5}	0.003

* Family wise error rate (*FWER*) cutoff of 0.1 was used to select DMRs

Table 3.6. DMRs of normal hematopoiesis located in super-enhancer of different tissues and cell types **(See Appendix3)**

Table 3.7. Common genes between mouse and human hematopoiesis (**See Appendix4**)

Columns are chromosome, start, end, diffMethyl (difference of methylation percentage for group2-group1 for group1 vs group2 comparison), p.value, fwer, Gene, annotation, relation to gene, distance to gene.

Yellow colored DMRs are showing opposite direction of methylation change compared to mouse hematopoietic DMRs.

Table 3.8. Primers for bisulfite pyrosequencing

Gene	Primer type	Sequence(5'→3')
MIR539	Nested forward	TATGATAAGTTTTGTAAAGGGATGTA
	Nested reverse	/5Biosg/CAAATCCCTAATAACACCAAAAAAT
	Long forward	GTGTTGTTGTTTTATATTTGAGGAGAA
	Long reverse	CATATCCAAAAAATACCTCCAAAAA
	Sequencing 1 (F)	TGATAAGTTTTGTAAAGGGATG
	Sequencing 2 (F)	GTTTTAATTTTAGAATTTTGGA
CDK6	Nested forward	TGTTTTTGAGATAGTAGTAGGGTATTTTG
	Nested reverse	/5Biosg/TAACCAATCTAAACCCCATTTACTC
	Long forward	GGGGTAGATAGTTTTATATAGGGTAGTTGT
	Long reverse	TTCCACCCCAAAATTTATTATAACA
	Sequencing 1 (F)	GATAGTAGTAGGGTATTTTGAT
	Sequencing 2 (F)	ATTGTTTTTTTTTTGTTAAAGG
	Sequencing 3 (F)	TAAGTGGGAATTAAGTTTTGAG
SLC39A4	Nested forward	AGGGGAAGGTAGATTTTAGGGTAG
	Nested reverse	/5Biosg/AACCCCAAAACACTAAACTCAATAC
	Long forward	TTTTGAGTTTAGAGGTTTTATTTTTAT
	Long reverse	ACAAACTCCCCTAAAAACCC
	Sequencing 1 (F)	AGGGGGAGTTTAGATGTTATTT
	Sequencing 2 (F)	GTTTGAGGTTTAAGGATTTTGT
ZDHHC14	Nested forward	GGGAAAGAAGAGAATTATTTTTAGGTT
	Nested reverse	/5Biosg/CAAACCCAATACCTCTATCAAAATC
	Long forward	GTGTTAATGGTATTTTTTGATAGT
	Long reverse	AAACTATCTTTACTTTTACTCAAAC
	Sequencing 1 (F)	GAGGAAATGGGAGTTTTGTTTT
	Sequencing 2 (F)	GAGTTTATTTATGTTTGTAGAT
	Sequencing 3 (F)	AAATATATTTTTTTTTTTATT
	Sequencing 4 (F)	GAAGTTTTTTTTGATTTTTGTTT

TRPM2	Nested forward	GGGTTGTTTAGAAGGGTTATTGATT
	Nested reverse	/5Biosg/CCCAATTCTATTCTCCCAAAAATATA
	Long forward	TTTTTAGTTTTGAGGAAAGTTGGTT
	Long reverse	AATAATCATACAACCCACAAAAAAC
	Sequencing 1 (F)	GATTTGGGGATGGTTTTTAATT
	Sequencing 2 (F)	GAGTTGGAGGTTATAGTGTTTT
	Sequencing 3 (F)	TTGTTGGTTAGTTTGTAGTTGG
	Sequencing 4 (F)	TTGTTGGTTAGTTTGTAGTTGG
MAEA	Nested forward	TTAGTTTAGGATGGTAGGAAGTTAT
	Nested reverse	/5Biosg/TTATCTTTTACAATTAACCAAAAAA
	Long forward	TTTTTAAGAAGTTTTTTAGGGGATT
	Long reverse	ATATCATCATCATATTTTCACCAACAC
	Sequencing 1 (F)	TAGTTTAGGATGGTAGGAAGTT
	Sequencing 2 (F)	AATATGAGATTGGTTTTTTTAG
	Sequencing 3 (F)	TATTTTTGGTTTTGTGATGAGT
STAT3	Nested forward	GGTTTTTAATGTAGGTAATTTGTTGT
	Nested reverse	/5Biosg/TATTTTAATTTCCAACCAAAACATC
	Long forward	TTTTTTATTTTTTATTAGTTTTTTAGT
	Long reverse	CTTACTTAATTTCTAAAAAATTCCTACTCT
	Sequencing 1 (F)	TTTTAATGTAGGTAATTTGTT
	Sequencing 2 (F)	GTGAGAGTTTTTTG
HMHB1	Nested forward	TGGAGAAATTAGAATTGGAGGAGTA
	Nested reverse	/5Biosg/CTAAATAATCCCAACAACAAAAACC
	Long forward	ATGAGGAAATTATATTTTAGGAGGT
	Long reverse	CAACCAAACAATAAACTATAAAACC
	Sequencing 1 (F)	GAGAAGAAAAAAGAGGTGAGGG
	Sequencing 2 (F)	TATAATAGGTGAAAATAGGGAT
MPO	Nested forward	TAGTTTTAGTTGGTTGGATATGTTG
	Nested reverse	/5Biosg/AACCTCTCTCTATACCTCAAATCCC
	Long forward	TAGGTTGTTAAAGGGTAGTAGGGTT

Long reverse	TACCAAAAATCCTAAAAACAAAAAA
Sequencing 1 (F)	AGTTTTAGTTGGTTGGATATGT
Sequencing 2 (F)	GTAGGTTTTTGGTTAGGGGTTT
Sequencing 3 (F)	GGATGGTGATGTTGTT

/5Biosg/ = 5’ biotin added

F = forward

Chapter 4

The cell of origin of leukemia stem cell

This work is an ongoing project of the Feinberg Lab and Johns Hopkins. All publication rights are reserved for these institutions and the presentation of this work here does not preclude future publication elsewhere.

Summary

The question of the cell of origin of LSC has long been remained in the field, yet much of it still needs to be understood. We hypothesized that DNA methylation could be useful to infer the cell identity of LSCs compared to HSPCs, as it has been shown that transformed or reprogrammed cells retained an epigenetic memory of the cell of origin (Kim et al., 2010; Kim et al., 2011; Polak et al., 2015). We were able to provide a clue for the cell of origin of LSCs by comparing the DNA methylation profile of LSCs to normal HSPCs, obtained from previous chapter. Clustering analysis based on DNA methylation status showed LSCs clustered with either L-MPP or GMP. This result was replicated in TCGA data set, a larger cohort, and supported by molecular characteristics of each cluster such as FAB types, genetic mutations, and cytogenetic risk groups. These results suggested L-MPP and GMP would be major sources of LSCs.

Results

LSC Forms Two Clusters with L-MPP-like and GMP-like

In order to relate normal hematopoiesis to LSC, we first identified the DMRs from all possible pairwise comparisons among the 6 HSPC after applying a more rigorous cutoff of *FWER*<0.1 (Table 4.1). The resulting 216 DMRs were applied in clustering analysis including all 6 normal HSPC populations with LSCs and Blasts

(Figure 4.1). Strikingly, this analysis revealed that AML samples formed 2 distinct clusters, L-MPP-like and GMP-like (Figure 4.1). Importantly, the GMP-like cluster included several CD34+CD38- subpopulations, indicating that these clusters could not have been identified by immunophenotype alone. Moreover, clustering analysis using an equal number of length-matched random regions showed that the clustering of AML populations with either L-MPP or GMP was unique to the selected DMRs (Figure 4.2).

TCGA AML Samples Form Two Major Clusters, L-MPP-like and GMP-like

Strikingly, using the same 216 DMRs, the TCGA samples also formed the same two major clusters, L-MPP-like and GMP-like (Figure 4.3). In addition to the two major clusters, we also identified a minor CMP-like cluster that was not observed in our smaller cohort. We calculated scores indicating the similarity of each TCGA sample to each of the six progenitors, and designated a counterpart HSPC population for each TCGA sample based on highest similarity. This approach showed that 76.6% of TCGA samples resembled GMP and 14.6% had a methylation profile most similar to L-MPP (Figure 4.4).

We hypothesized that if the assignment of AML samples to L-MPP-like and GMP-like clusters was related to the cell of origin, then the degree of maturity and morphology might differ between the two groups. Consistent with this, we compared the distribution of the French-American-British (FAB) classification of the TCGA samples and found that the L-MPP-like cases mainly consisted of more immature M0, M1, and M2 types, while the more differentiated M4 and M5 types were enriched in GMP-like AML ($p<1x10^{-4}$, chi-square test, Figure 4.5). It should be noted that the LSC epigenetic signature is not merely a recapitulation of FAB types, as our signature is prognostic in multivariate analysis while FAB types are not. In addition, it is not possible to know the

cell of origin simply by examining FAB types (Table 4.2). For example, in 38 cases of M1 AML, 29/38 are GMP-like and 9/29 are L-MPP-like (Table 4.2).

Finally, we sought to investigate if the L-MPP-like and GMP-like clusters, and therefore the potential cell of origin, were associated with cytogenetic abnormalities or recurrent mutations of specific genes including *DNMT3A, IDH1, IDH2, TET1, TET2, FLT3*, and *NPM1*. The GMP-like cluster was enriched for patients in the low and intermediate cytogenetic risk groups, while the L-MPP-like cluster was enriched for patients in the high cytogenetic risk group (p=1x10^{-4}, Fisher's exact test; Figure 4.6a). We found that *IDH1* and *IDH2* mutations were enriched in the L-MPP like group (p<0.01 for both, Fisher's exact test), and *FLT3* and *NPM1* mutations were enriched in the GMP-like group (p<0.01 for both, Fisher's exact test). *DNMT3A* and *TET1* mutations were more enriched in the L-MPP group, but this was not statistically significant (Figure 4.6b). Together, these results demonstrate that DNA methylation signatures permit a novel clustering of AML into L-MPP-like and GMP-like groups that may reflect the cell of origin for each case and demonstrate an association with key disease features.

Discussion

As DNA methylation is a potential marker of cell identity, here we compared DNA methylomes of normal HSPC to LSC as a marker for the cell of origin in AML. Using this approach, we observed two subtypes in our cohort: L-MPP-like and GMP-like. These two subtypes were also identified in the TCGA cohort, suggesting that leukemic transformation predominantly occurs at either the L-MPP or GMP stage of hematopoietic development. We found that other features of AML were associated with these two

subtypes including FAB type, several mutations, and cytogenetic abnormalities, suggesting that the cell of origin may drive key clinical features in AML.

We newly identified a small subset of AML cells clustering with CMP and few samples clustering with HSC, MPP, and MEP that could not be identified in smaller datasets including our own data and a previous study (Goardon et al., 2011). The result from TCGA indicates that the cell of origin of AML could be variable among HSPCs. The result for the cell of origin should be considered carefully, as it is hard to provide definite proof of the question due to experimental limitations. Despite this caveat that affects all of human primary cancer biology, we do provide the first epigenetic evidence for cell of origin in human leukemia and believe that our approach using epigenomic profiles suggests an efficient way to study cell of origin in cancer biology using large data sets.

Materials and Methods

Illumina Infinium Human Methylation 450 Bead Array Assay

Genomic DNA from each sample was purified using the MasterPure DNA purification kit (Epicentre) according to the manufacturer's protocol. The genomic DNA (250-500ng) was treated with sodium bisulfate using the Zymo EZ DNA Methylation Kit (ZYMO Research) as recommended by the manufacturer, with the alternative incubation conditions for the Illumina Infinium Methylation Assay. Converted DNA was eluted in 11ul of elution buffer. DNA methylation level was measured using Illumina Infinium HD Methylation Assay (Illumina) according to the manufacturer's specifications.

Methylation array data are deposited at the Gene Expression Omnibus (GEO) with accession number GSE63409.

Illumina Infinium Human Methylation 450 Bead Array Analysis

Raw intensity files were obtained using minfi package (Aryee et al., 2014) to calculate methylation ratios (Beta values). The data was normalized using Illumina preprocessing method implemented in minfi. Several quality control measures were applied to remove arrays with low quality. Control probes were examined on the 450k array to assess several measures including bisulfite conversion, extension, hybridization, specificity and others. One of the MPP samples (BM2712) and two samples of TCGA (Patient ID: 2934 and 2827) showed low quality for the measures, so they were removed for further analysis. Next, median methylated and unmethylated signals were calculated for each arrays; no array was identified for signal values lower than 10.5. For multidimensional scaling analysis, probes containing an annotated SNP (dbSNP137) at the single-base extension or CpG sites were removed (17398 probes removed). Minfi 1.8.9 was used.

Bump hunting method previously described was applied to identify DMRs in 450k array (Aryee et al., 2014; Jaffe et al., 2012). Beta value of 0.1 (10% of methylation difference) was used as cutoff when finding DMRs. Bumphunter 1.2.0 was used.

Statistical Analysis

To assign cell identity of normal HSPCs to TCGA samples, mean methylation value of each 216 DMRs for normal hematopoiesis (methylation profile) was retrieved and standard deviation of the mean value for each signature was calculated. Then scores (probability density values as log value) for each TCGA sample regarding normal HSPCs' profile was calculated using dnorm function with the mean and standard deviation

calculated in previous step. Maximum value of scores among the ones regarding normal HSPC methylation profile was chosen, and then cell identity assigned.

For clustering analysis, hclust function with ward method in R was used to generate all the cluster dendrogram analysis.

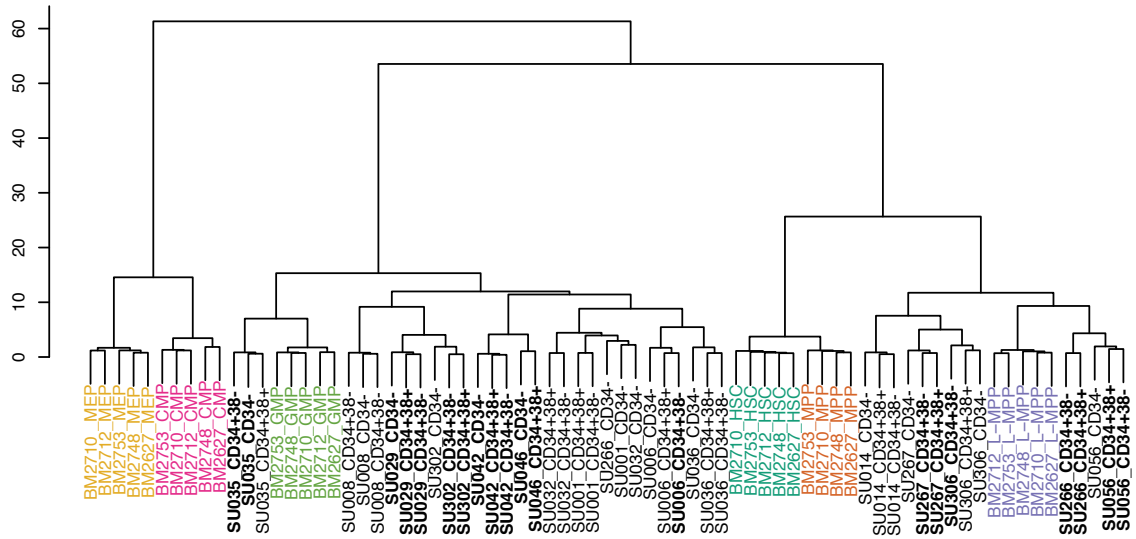


Figure 4.1. Epigenetic signatures define subgroups of AML LSC reflecting the cell of origin. A total of 216 DMRs identified from all possible pairwise comparisons among 6 HSPCs were used to cluster all normal HSPC with all AML subpopulations. The primary AML subpopulations form two major clusters: L-MPP-like and GMP-like. LSC subpopulations are indicated in bold.

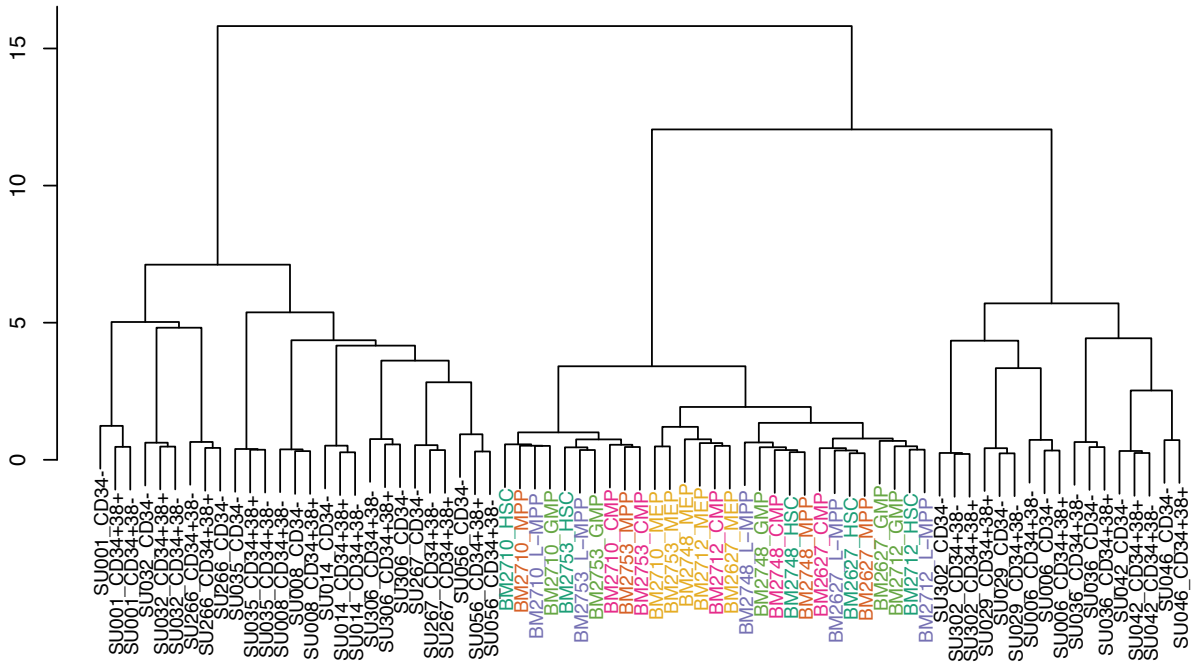


Figure 4.2. Clustering analysis of AML populations with normal HSPCs using length matched random 216 regions. Clustering analysis using random length matched regions shows no clustering between AML populations or normal HSPCs. Normal progenitors clustered together, but not by lineages.

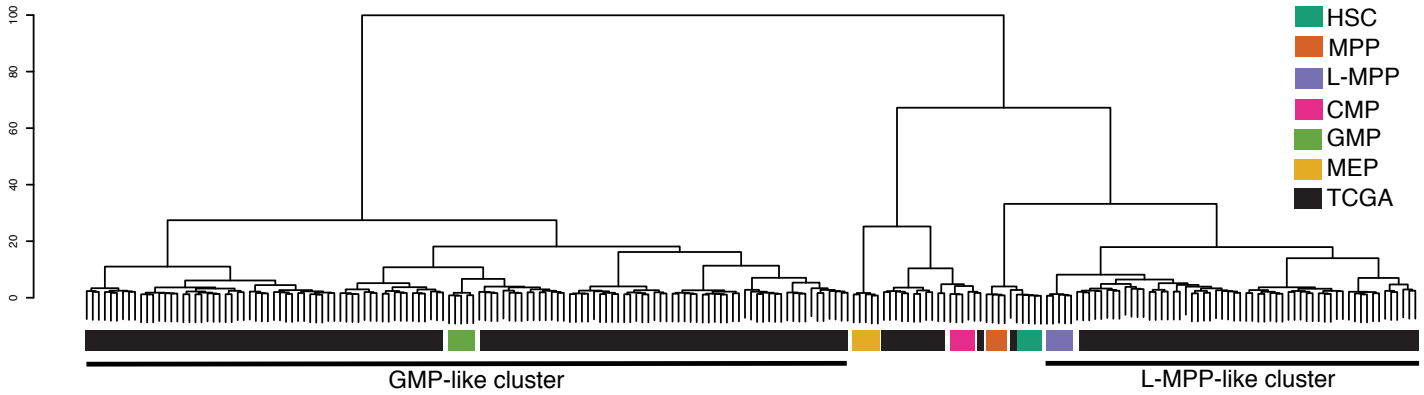


Figure 4.3. Epigenetic signatures define subgroups of AML samples in TCGA reflecting the cell of origin. Clustering analysis of TCGA AML samples with normal human HSPC using the 216 DMRs shows that the L-MPP-like and GMP-like clusters are observed in this cohort as well.

Cell type	TCGA	%
HSC	1	0.5
MPP	3	1.6
L-MPP	28	14.6
CMP	12	6.3
GMP	147	76.6
MEP	1	0.5
	192	100

Figure 4.4. Cell identity of TCGA AML samples. TCGA samples were classified into one of HSPC populations according to their DNA methylation profile by generating methylation profiles of all the normal HSPC, and then calculating scores of each sample based on the closest population. The normal progenitor cell identity for each TCGA samples is summarized in this table.

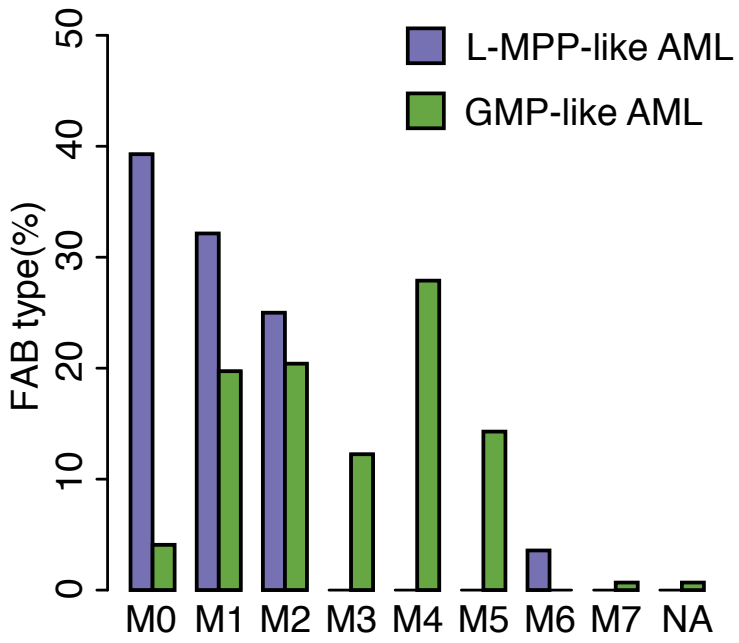


Figure 4.5. Distribution of FAB types for L-MPP-like and GMP-like TCGA AML samples. The L-MPP-like and GMP-like TCGA samples were grouped according to their FAB classification. NA: not classified.

a

Cell identity	Favorable	Intermediate/Normal	Poor
GMP-like	33	90	22
L-MPP-like	0	13	15

b

	L-MPP (%)	GMP (%)
DNMT3A	33.3	25.0
IDH1	33.3	6.3
IDH2	29.6	5.6
TET1	3.7	0.7
TET2	7.4	7.6
NPM1	3.7	34.7
FLT3	7.4	32.6

Figure 4.6. Correlation of disease features with cell identity. (**a**) Number of TCGA samples that belong to each cytogenetic risk group is shown for GMP-like and L-MPP-like AML samples. Information on cytogenetic risk group was retrieved from clinical annotation of TCGA patients. (**b**) Percentage of AML samples with specific mutations including *DNMT3A*, *IDH1*, *IDH2*, *TET1*, *TET2*, *NPM1,* and *FLT3* are shown for L-MPP-like and GMP-like AML samples. Mutations that showed statistical significance of association with specific cell types are colored in red. Information on mutations was retrieved from TCGA Mutation Annotation Format (MAF) file.

Table 4.1. DMRs for normal hematopoiesis

chr	start	end	gene name
chr1	2164542	2164602	SKI
chr1	12538341	12538541	VPS13D
chr1	19652788	19652788	PQLC2
chr1	25598878	25599066	RHD
chr1	25747376	25747437	RHCE
chr1	39513326	39513326	NDUFS5
chr1	43290834	43291209	ERMAP
chr1	44114346	44114355	KDM4A
chr1	45103896	45103967	RNF220
chr1	51986130	51986254	EPS15
chr1	53103294	53103294	FAM159A
chr1	116193783	116193783	VANGL1
chr1	145828001	145828006	GPR89A
chr1	159866558	159866558	CCDC19
chr1	179016233	179016233	FAM20B
chr1	182760143	182761262	NPL
chr1	185253495	185253495	SWT1
chr1	199827580	199827580	NR5A2
chr1	226012913	226013010	EPHX1
chr1	226036279	226036279	TMEM63A
chr10	11631072	11631072	USP6NL
chr10	32621403	32621403	EPC1
chr10	70805603	70805603	KIAA1279
chr10	97036527	97036527	PDLIM1
chr10	100220809	100220809	HPSE2
chr10	135059097	135059289	MIR202

chr11	1297066	1297087	TOLLIP
chr11	1325718	1325852	TOLLIP
chr11	5276490	5276490	HBG2
chr11	8385712	8385767	STK33
chr11	12136405	12136405	MICAL2
chr11	33759043	33759043	CD59
chr11	47885166	47885166	NUP160
chr11	59823993	59824161	MS4A3
chr11	67251677	67251939	AIP
chr11	69061454	69061473	MYEOV
chr11	73681281	73681281	DNAJB13
chr11	108422791	108423124	EXPH5
chr11	117695591	117696015	FXYD2
chr12	1922058	1922067	CACNA2D4
chr12	14996508	14996587	ART4
chr12	25103173	25103643	BCAT1
chr12	32530696	32530696	BICD1
chr12	53602179	53602179	ITGB7
chr12	89938002	89938002	POC1B-GALNT4
chr12	96390059	96390143	HAL
chr12	114232702	114232905	RBM19
chr12	117480333	117480333	TESC
chr12	123632825	123633306	PITPNM2
chr13	41631052	41631052	WBP4
chr13	49147573	49147573	RCBTB2
chr13	113305704	113305901	C13orf35
chr13	114828264	114828455	RASA3
chr13	114918702	114918702	RASA3
chr14	21359295	21359943	RNASE3

chr14	81425912	81426015	TSHR
chr14	101513572	101514051	MIR539
chr14	103565867	103566172	EXOC3L4
chr14	104190678	104190829	ZFYVE21
chr14	104197160	104197160	ZFYVE21
chr14	104625249	104625669	KIF26A
chr15	60907749	60907749	RORA
chr15	65066231	65066710	RBPMS2
chr15	80938098	80938098	ARNT2
chr15	90643766	90643766	IDH2
chr15	90727560	90727570	SEMA4B
chr15	101720914	101720914	CHSY1
chr15	101777761	101777800	CHSY1
chr16	23880664	23880664	PRKCB
chr16	56892460	56892460	MIR138-2
chr16	67567158	67567158	FAM65A
chr16	85551478	85551748	KIAA0182
chr16	85622205	85622276	KIAA0182
chr16	85708039	85708039	KIAA0182
chr16	86011615	86012085	IRF8
chr16	88558065	88558237	ZFPM1
chr16	88563934	88564934	ZFPM1
chr17	998432	998504	ABR
chr17	5138259	5138696	LOC100130950
chr17	14222235	14222235	HS3ST3B1
chr17	26517057	26517057	NLK
chr17	38717206	38717275	CCR7
chr17	40489569	40489785	STAT3
chr17	45286354	45286609	MYL4

chr17	46648525	46648582	HOXB3
chr17	47372126	47372126	ZNF652
chr17	74844962	74845007	MGAT5B
chr17	76142333	76142461	C17orf99
chr17	76850256	76850277	TIMP2
chr17	79623603	79623661	PDE6G
chr17	80273242	80273322	CD7
chr17	80581701	80581925	WDR45L
chr18	55250568	55250568	FECH
chr18	60646614	60646671	PHLPP1
chr19	827715	827843	AZU1
chr19	2865997	2865997	ZNF556
chr19	12895288	12895529	JUNB
chr19	12997997	12998927	KLF1
chr19	15391832	15391946	BRD4
chr19	39048014	39048014	RYR1
chr19	47839132	47839187	GPR77
chr19	50848024	50848461	NR1H2
chr19	51858208	51858276	ETFB
chr19	52076691	52076691	ZNF175
chr19	55417390	55417647	NCR1
chr2	24233923	24234017	MFSD2B
chr2	37642493	37642493	QPCT
chr2	39355435	39355435	SOS1
chr2	55642894	55642894	CCDC88A
chr2	61407270	61407414	AHSA2
chr2	65593761	65594021	SPRED2
chr2	68348796	68348796	WDR92
chr2	74753691	74753759	AUP1

chr2	75089669	75089669	HK2
chr2	109789810	109789810	SH3RF3
chr2	113885116	113885277	IL1RN
chr2	128052778	128052889	ERCC3
chr2	144992352	144992352	GTDC1
chr2	149259371	149259371	MBD5
chr2	149639612	149639914	KIF5C
chr2	170779769	170779769	UBR3
chr2	172377981	172378036	CYBRD1
chr2	179149472	179149605	OSBPL6
chr2	220436894	220436948	INHA
chr2	239330171	239330202	ASB1
chr2	240225062	240225142	HDAC4
chr20	45319350	45319455	TP53RK
chr20	52199594	52199778	ZNF217
chr21	43823749	43824262	UBASH3A
chr21	45773189	45774294	TRPM2
chr22	24108000	24108140	CHCHD10
chr22	39106452	39106452	GTPBP1
chr3	3123308	3123308	IL5RA
chr3	12994821	12994919	IQSEC1
chr3	46394550	46395356	CCR2
chr3	49127364	49127364	QRICH1
chr3	58200471	58200569	DNASE1L3
chr3	71295684	71295684	FOXP1
chr3	127473936	127473936	MGLL
chr3	128998952	128998952	C3orf37
chr3	148581801	148581837	CPA3
chr3	156537970	156537970	LOC730091

chr3	176919597	176919597	TBL1XR1
chr3	184297380	184297522	EPHB3
chr3	193405999	193405999	OPA1
chr3	195603697	195603697	TNK2
chr3	196351986	196352142	LRRC33
chr4	1195845	1196179	SPON2
chr4	1294783	1295078	MAEA
chr4	2431872	2432313	LOC402160
chr4	3374767	3374909	RGS12
chr4	24590014	24590014	DHX15
chr4	38859706	38859770	TLR6
chr4	87938832	87938832	AFF1
chr4	144266241	144266241	GAB1
chr4	157897045	157897045	PDGFC
chr5	1089722	1089800	SLC12A7
chr5	1555791	1555887	SDHAP3
chr5	32297979	32297979	MTMR12
chr5	55148493	55148534	IL31RA
chr5	67586170	67586258	PIK3R1
chr5	125694354	125694418	GRAMD3
chr5	137224213	137224284	MYOT
chr5	143191565	143192067	HMHB1
chr5	153753045	153753045	GALNT10
chr5	169617918	169617918	C5orf58
chr5	177913434	177913485	COL23A1
chr6	6588693	6589075	LY86
chr6	16933816	16933816	FLJ23152
chr6	26030329	26030329	HIST1H3B
chr6	28885444	28885568	TRIM27

chr6	30956397	30956440	MUC21
chr6	31088343	31088434	PSORS1C1
chr6	31629088	31629199	GPANK1
chr6	31654389	31654533	ABHD16A
chr6	32099450	32099564	FKBPL
chr6	32121055	32121393	PPT2
chr6	32135715	32136052	EGFL8
chr6	32810742	32810833	PSMB8
chr6	32825040	32825897	PSMB9
chr6	32905085	32905320	HLA-DMB
chr6	41010111	41010316	UNC5CL
chr6	41254433	41254471	TREM1
chr6	109812795	109812795	AKD1
chr6	135517041	135517046	MYB
chr6	140168822	140168822	LOC100132735
chr6	142721804	142721804	GPR126
chr6	157876915	157877120	ZDHHC14
chr7	1545385	1545819	INTS1
chr7	2653651	2654120	IQCE
chr7	30108301	30108301	PLEKHA8
chr7	33765409	33765409	BBS9
chr7	65419185	65419288	VKORC1L1
chr7	73645723	73645783	RFC2
chr7	80267619	80267943	CD36
chr7	101361395	101361745	MYL10
chr7	138347826	138348384	SVOPL
chr7	138816336	138816336	TTC26
chr7	142659349	142659425	KEL
chr8	1708438	1708526	CLN8

chr8	1870722	1870798	ARHGEF10
chr8	12608579	12608579	LONRF1
chr8	13105155	13105155	DLC1
chr8	33421410	33421410	RNF122
chr8	41654331	41655078	ANK1
chr8	42125496	42125496	IKBKB
chr8	42623730	42623946	CHRNA6
chr8	68022262	68022262	CSPP1
chr8	99105576	99105576	C8orf47
chr8	128972450	128972829	PVT1
chr8	131368433	131368433	ASAP1
chr8	141312892	141312979	TRAPPC9
chr8	145643083	145643626	SLC39A4
chr9	132145105	132145105	C9orf106
chr9	136726359	136726575	VAV2
chr9	137277819	137277819	RXRA

Table 4.2. FAB type distribution for L-MPP-like and GMP-like AML samples

	M0	M1	M2	M3	M4	M5	M6	M7	NA
L-MPP	11	9	7	0	0	0	1	0	0
GMP	6	29	30	18	41	21	0	1	1

References

Abdel-Wahab, O., Adli, M., LaFave, L.M., Gao, J., Hricik, T., Shih, A.H., Pandey, S., Patel, J.P., Chung, Y.R., Koche, R.*, et al.* (2012). ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. Cancer cell *22*, 180-193.

Alharbi, R.A., Pettengell, R., Pandha, H.S., and Morgan, R. (2013). The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia *27*, 1000-1008.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363-1369.

Bach, C., Buhl, S., Mueller, D., Garcia-Cuellar, M.P., Maethner, E., and Slany, R.K. (2010). Leukemogenic transformation by HOXA cluster genes. Blood *115*, 2910-2918.

Bartholdy, B., Christopeit, M., Will, B., Mo, Y., Barreyro, L., Yu, Y., Bhagat, T.D., Okoye-Okafor, U.C., Todorova, T.I., Greally, J.M.*, et al.* (2014). HSC commitment-associated epigenetic signature is prognostic in acute myeloid leukemia. The Journal of clinical investigation *124*, 1158-1167.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell *143*, 470-484.

Bassil, C.F., Huang, Z., and Murphy, S.K. (2013). Bisulfite pyrosequencing. Methods in molecular biology *1049*, 95-107.

Becker, A.J., Mc, C.E., and Till, J.E. (1963). Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature *197*, 452-454.

Benetatos, L., Hatzimichael, E., Londin, E., Vartholomatos, G., Loher, P., Rigoutsos, I., and Briasoulis, E. (2013). The microRNAs within the DLK1-DIO3 genomic region: involvement in disease pathogenesis. Cellular and molecular life sciences : CMLS *70*, 795-814.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R., and Sultan, C. (1976). Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. British journal of haematology *33*, 451-458.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R., and Sultan, C. (1985). Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group. Annals of internal medicine *103*, 620-625.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R.*, et al.* (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology *28*, 1045-1048.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.

Bock, C., Beerman, I., Lien, W.H., Smith, Z.D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D.J., and Meissner, A. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Molecular cell *47*, 633-647.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature medicine *3*, 730-737.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. Nature *371*, 435-438.

Brinkman, A.B., Gu, H., Bartels, S.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A.*, et al.* (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome research *22*, 1128-1138.

Broske, A.M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R.*, et al.* (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nature genetics *41*, 1207-1215.

Bullinger, L., Ehrich, M., Dohner, K., Schlenk, R.F., Dohner, H., Nelson, M.R., and van den Boom, D. (2010). Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. Blood *115*, 636-642.

Busque, L., Patel, J.P., Figueroa, M.E., Vasanthakumar, A., Provost, S., Hamilou, Z., Mollica, L., Li, J., Viale, A., Heguy, A.*, et al.* (2012). Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nature genetics *44*, 1179-1181.

Byrd, J.C., Mrozek, K., Dodge, R.K., Carroll, A.J., Edwards, C.G., Arthur, D.C., Pettenati, M.J., Patil, S.R., Rao, K.W., Watson, M.S.*, et al.* (2002). Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). Blood *100*, 4325-4336.

Cancer Genome Atlas Research, N. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. The New England journal of medicine *368*, 2059-2074.

Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y.*, et al.* (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nature genetics *44*, 23-31.

Chao, M.P., Seita, J., and Weissman, I.L. (2008). Establishment of a normal hematopoietic and leukemia stem cell hierarchy. Cold Spring Harbor symposia on quantitative biology *73*, 439-449.

Chung, K.Y., Morrone, G., Schuringa, J.J., Plasilova, M., Shieh, J.H., Zhang, Y., Zhou, P., and Moore, M.A. (2006). Enforced expression of NUP98-HOXA9 in human CD34(+) cells enhances stem cell proliferation. Cancer Res *66*, 11781-11791.

Civin, C.I., Strauss, L.C., Brovall, C., Fackler, M.J., Schwartz, J.F., and Shaper, J.H. (1984). Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. Journal of immunology *133*, 157-165.

Clark, S.J., Harrison, J., Paul, C.L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. Nucleic acids research *22*, 2990-2997.

Conerly, M.L., Teves, S.S., Diolaiti, D., Ulrich, M., Eisenman, R.N., and Henikoff, S. (2010). Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. Genome research *20*, 1383-1390.

Corces-Zimmerman, M.R., Hong, W.J., Weissman, I.L., Medeiros, B.C., and Majeti, R. (2014). Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proceedings of the National Academy of Sciences of the United States of America *111*, 2548-2553.

Cortazar, D., Kunz, C., Saito, Y., Steinacher, R., and Schar, P. (2007). The enigmatic thymine DNA glycosylase. DNA repair *6*, 489-504.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D.*, et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67-79.

Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Basavapathruni, A., Jin, L., Boriack-Sjodin, P.A., Allain, C.J., Klaus, C.R., Raimondi, A., Scott, M.P.*, et al.* (2013). Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood *122*, 1017-1025.

Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Majer, C.R., Sneeringer, C.J., Song, J., Johnston, L.D., Scott, M.P., Smith, J.J., Xiao, Y.*, et al.* (2011). Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. Cancer cell *20*, 53-65.

Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C.*, et al.* (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature *462*, 739-744.

de Jonge, H.J., Woolthuis, C.M., Vos, A.Z., Mulder, A., van den Berg, E., Kluin, P.M., van der Weide, K., de Bont, E.S., Huls, G., Vellenga, E.*, et al.* (2011). Gene expression profiling in the leukemic stem cell-enriched CD34+ fraction identifies target genes that predict prognosis in normal karyotype AML. Leukemia *25*, 1825-1833.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes & development *25*, 1010-1022.

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., and Fuks, F. (2011). Evaluation of the Infinium Methylation 450K technology. Epigenomics *3*, 771-784.

Deneberg, S., Guardiola, P., Lennartsson, A., Qu, Y., Gaidzik, V., Blanchet, O., Karimi, M., Bengtzen, S., Nahi, H., Uggla, B.*, et al.* (2011). Prognostic DNA methylation

patterns in cytogenetically normal acute myeloid leukemia are predefined by stem cell chromatin marks. Blood *118*, 5573-5582.

DiGiusto, D., Chen, S., Combs, J., Webb, S., Namikawa, R., Tsukamoto, A., Chen, B.P., and Galy, A.H. (1994). Human fetal bone marrow early progenitors for T, B, and myeloid cells are found exclusively in the population expressing high levels of CD34. Blood *84*, 421-432.

Dohner, H. (2007). Implication of the molecular characterization of acute myeloid leukemia. Hematology / the Education Program of the American Society of Hematology American Society of Hematology Education Program, 412-419.

Doi, A., Park, I.H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S.*, et al.* (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nature genetics *41*, 1350-1353.

Doulatov, S., Notta, F., Eppert, K., Nguyen, L.T., Ohashi, P.S., and Dick, J.E. (2010). Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nature immunology *11*, 585-593.

Doulatov, S., Notta, F., Laurenti, E., and Dick, J.E. (2012). Hematopoiesis: a human perspective. Cell stem cell *10*, 120-136.

Drabkin, H.A., Parsy, C., Ferguson, K., Guilhot, F., Lacotte, L., Roy, L., Zeng, C., Baron, A., Hunger, S.P., Varella-Garcia, M.*, et al.* (2002). Quantitative HOX expression in chromosomally defined subsets of acute myelogenous leukemia. Leukemia *16*, 186-195.

Dufour, A., Schneider, F., Metzeler, K.H., Hoster, E., Schneider, S., Zellmeier, E., Benthaus, T., Sauerland, M.C., Berdel, W.E., Buchner, T.*, et al.* (2010). Acute myeloid leukemia with biallelic CEBPA gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome. J Clin Oncol *28*, 570-577.

Eppert, K., Takenaka, K., Lechman, E.R., Waldron, L., Nilsson, B., van Galen, P., Metzeler, K.H., Poeppl, A., Ling, V., Beyene, J.*, et al.* (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. Nature medicine *17*, 1086-1093.

Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A.V., Waghorn, K., Zoi, K., Ross, F.M., Reiter, A.*, et al.* (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nature genetics *42*, 722-726.

Estey, E.H. (2013). Epigenetics in clinical practice: the examples of azacitidine and decitabine in myelodysplasia and acute myeloid leukemia. Leukemia *27*, 1803-1812.

Feinberg, A.P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. Nature reviews Genetics *7*, 21-33.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F.*, et al.* (2010a). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer cell *18*, 553-567.

Figueroa, M.E., Lugthart, S., Li, Y., Erpelinck-Verschueren, C., Deng, X., Christos, P.J., Schifano, E., Booth, J., van Putten, W., Skrabanek, L.*, et al.* (2010b). DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer cell *17*, 13-27.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences of the United States of America *89*, 1827-1831.

Genovese, G., Kahler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M.*, et al.* (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. The New England journal of medicine *371*, 2477-2487.

Gentles, A.J., Plevritis, S.K., Majeti, R., and Alizadeh, A.A. (2010). Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. Jama *304*, 2706-2715.

Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K.A., Rout, R.*, et al.* (2011). Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. Cancer cell *19*, 138-152.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A*., et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531-537.

Grimwade, D., Walker, H., Oliver, F., Wheatley, K., Harrison, C., Harrison, G., Rees, J., Hann, I., Stevens, R., Burnett, A*., et al.* (1998). The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. Blood *92*, 2322-2333.

Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome biology *13*, R83.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D*., et al.* (2011). Increased methylation variation in epigenetic domains across cancer types. Nature genetics *43*, 768-775.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L*., et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Herb, B.R., Wolschin, F., Hansen, K.D., Aryee, M.J., Langmead, B., Irizarry, R., Amdam, G.V., and Feinberg, A.P. (2012). Reversible switching between epigenetic states in honeybee behavioral subcastes. Nature neuroscience *15*, 1371-1373.

Heuser, M., Yun, H., Berg, T., Yung, E., Argiropoulos, B., Kuchenbauer, F., Park, G., Hamwi, I., Palmqvist, L., Lai, C.K*., et al.* (2011). Cell of origin in AML: susceptibility to MN1-induced transformation is regulated by the MEIS1/AbdB-like HOX protein complex. Cancer cell *20*, 39-52.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934-947.

Hosen, N., Park, C.Y., Tatsumi, N., Oji, Y., Sugiyama, H., Gramatzki, M., Krensky, A.M., and Weissman, I.L. (2007). CD96 is a leukemic stem cell-specific marker in human acute myeloid leukemia. Proceedings of the National Academy of Sciences of the United States of America *104*, 11008-11013.

Huang, Y., Pastor, W.A., Shen, Y., Tahiliani, M., Liu, D.R., and Rao, A. (2010). The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PloS one *5*, e8888.

Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddeloh, J.A., Wen, B., and Feinberg, A.P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). Genome research *18*, 780-790.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M.*, et al.* (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature genetics *41*, 178-186.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.

Jaffe, A.E., Murakami, P., Lee, H., Leek, J.T., Fallin, M.D., Feinberg, A.P., and Irizarry, R.A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol *41*, 200-209.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A.*, et al.* (2014). Age-related clonal hematopoiesis associated with adverse outcomes. The New England journal of medicine *371*, 2488-2498.

Jamieson, C.H., Ailles, L.E., Dylla, S.J., Muijtjens, M., Jones, C., Zehnder, J.L., Gotlib, J., Li, K., Manz, M.G., Keating, A.*, et al.* (2004). Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. The New England journal of medicine *351*, 657-667.

Jan, M., Chao, M.P., Cha, A.C., Alizadeh, A.A., Gentles, A.J., Weissman, I.L., and Majeti, R. (2011). Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker. Proceedings of the National Academy of Sciences of the United States of America *108*, 5009-5014.

Jan, M., Snyder, T.M., Corces-Zimmerman, M.R., Vyas, P., Weissman, I.L., Quake, S.R., and Majeti, R. (2012). Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. Science translational medicine *4*, 149ra118.

Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K.*, et al.* (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature *467*, 338-342.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews Genetics *13*, 484-492.

Kikushige, Y., Shima, T., Takayanagi, S., Urata, S., Miyamoto, T., Iwasaki, H., Takenaka, K., Teshima, T., Tanaka, T., Inagaki, Y.*, et al.* (2010). TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. Cell stem cell *7*, 708-717.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.*, et al.* (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285-290.

Kim, K., Zhao, R., Doi, A., Ng, K., Unternaehrer, J., Cahan, P., Huo, H., Loh, Y.H., Aryee, M.J., Lensch, M.W.*, et al.* (2011). Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. Nature biotechnology *29*, 1117-1119.

Klebanoff, S.J. (2005). Myeloperoxidase: friend and foe. Journal of leukocyte biology *77*, 598-625.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R.*, et al.* (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839-843.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

Kozar, K., and Sicinski, P. (2005). Cell cycle progression without cyclin D-CDK4 and cyclin D-CDK6 complexes. Cell cycle *4*, 388-391.

Krause, D.S., Fackler, M.J., Civin, C.I., and May, W.S. (1996). CD34: structure, biology, and clinical utility. Blood *87*, 1-13.

Kreso, A., and Dick, J.E. (2014). Evolution of the cancer stem cell model. Cell stem cell *14*, 275-291.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929-930.

Krivtsov, A.V., Figueroa, M.E., Sinha, A.U., Stubbs, M.C., Feng, Z., Valk, P.J., Delwel, R., Dohner, K., Bullinger, L., Kung, A.L.*, et al.* (2013). Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. Leukemia *27*, 852-860.

Kulis, M., Heath, S., Bibikova, M., Queiros, A.C., Navarro, A., Clot, G., Martinez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M.*, et al.* (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nature genetics *44*, 1236-1242.

Kumar, R., DiMenna, L., Schrode, N., Liu, T.C., Franck, P., Munoz-Descalzo, S., Hadjantonakis, A.K., Zarrin, A.A., Chaudhuri, J., Elemento, O.*, et al.* (2013). AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. Nature *500*, 89-92.

Laird, P.W. (2010). Principles and challenges of genomewide DNA methylation analysis. Nature reviews Genetics *11*, 191-203.

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M.A., and Dick, J.E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature *367*, 645-648.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J.*, et al.* (2010). Dynamic changes in the human methylome during differentiation. Genome research *20*, 320-331.

Lehnertz, B., Pabst, C., Su, L., Miller, M., Liu, F., Yi, L., Zhang, R., Krosl, J., Yung, E., Kirschner, J.*, et al.* (2014). The methyltransferase G9a regulates HoxA9-dependent transcription in AML. Genes Dev *28*, 317-327.

Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865-873.

Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J.*, et al.* (2010). DNMT3A mutations in acute myeloid leukemia. The New England journal of medicine *363*, 2424-2433.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D.*, et al.* (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Losada, A. (2014). Cohesin in cancer: chromosome segregation and beyond. Nat Rev Cancer *14*, 389-393.

Lowenberg, B., Downing, J.R., and Burnett, A. (1999). Acute myeloid leukemia. The New England journal of medicine *341*, 1051-1062.

Lund, K., Adams, P.D., and Copland, M. (2014). EZH2 in normal and malignant hematopoiesis. Leukemia *28*, 44-49.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes & development *8*, 2282-2292.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. The Journal of biological chemistry *286*, 35334-35338.

Majeti, R., Becker, M.W., Tian, Q., Lee, T.L., Yan, X., Liu, R., Chiang, J.H., Hood, L., Clarke, M.F., and Weissman, I.L. (2009a). Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. Proceedings of the National Academy of Sciences of the United States of America *106*, 3396-3401.

Majeti, R., Chao, M.P., Alizadeh, A.A., Pang, W.W., Jaiswal, S., Gibbs, K.D., Jr., van Rooijen, N., and Weissman, I.L. (2009b). CD47 is an adverse prognostic factor and

therapeutic antibody target on human acute myeloid leukemia stem cells. Cell *138*, 286-299.

Majeti, R., Park, C.Y., and Weissman, I.L. (2007). Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. Cell stem cell *1*, 635-645.

Malumbres, M., Sotillo, R., Santamaria, D., Galan, J., Cerezo, A., Ortega, S., Dubus, P., and Barbacid, M. (2004). Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. Cell *118*, 493-504.

Marcucci, G., Maharry, K., Wu, Y.Z., Radmacher, M.D., Mrozek, K., Margeson, D., Holland, K.B., Whitman, S.P., Becker, H., Schwind, S.*, et al.* (2010). IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *28*, 2348-2355.

Mardis, E.R., Ding, L., Dooling, D.J., Larson, D.E., McLellan, M.D., Chen, K., Koboldt, D.C., Fulton, R.S., Delehaunty, K.D., McGrath, S.D.*, et al.* (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. The New England journal of medicine *361*, 1058-1066.

Martelli, M.P., Pettirossi, V., Thiede, C., Bonifacio, E., Mezzasoma, F., Cecchini, D., Pacini, R., Tabarrini, A., Ciurnelli, R., Gionfriddo, I.*, et al.* (2010). CD34+ cells from AML with mutated NPM1 harbor cytoplasmic mutated nucleophosmin and generate leukemia in immunocompromised mice. Blood *116*, 3907-3922.

Mestas, J., and Hughes, C.C. (2004). Of mice and not men: differences between mouse and human immunology. Journal of immunology *172*, 2731-2738.

Metzeler, K.H., Hummel, M., Bloomfield, C.D., Spiekermann, K., Braess, J., Sauerland, M.C., Heinecke, A., Radmacher, M., Marcucci, G., Whitman, S.P.*, et al.* (2008). An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood *112*, 4193-4201.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol Cell *10*, 1107-1117.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X.*, et al.* (2011). Tet2 loss leads to

increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer cell *20*, 11-24.

Nakamura, T., Largaespada, D.A., Lee, M.P., Johnson, L.A., Ohyashiki, K., Toyama, K., Chen, S.J., Willman, C.L., Chen, I.M., Feinberg, A.P.*, et al.* (1996). Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. Nat Genet *12*, 154-158.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., and Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO reports *10*, 1235-1241.

Patel, J.P., Gonen, M., Figueroa, M.E., Fernandez, H., Sun, Z., Racevskis, J., Van Vlierberghe, P., Dolgalev, I., Thomas, S., Aminova, O.*, et al.* (2012). Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. The New England journal of medicine *366*, 1079-1089.

Polak, P., Karlic, R., Koren, A., Thurman, R., Sandstrom, R., Lawrence, M.S., Reynolds, A., Rynes, E., Vlahovicek, K., Stamatoyannopoulos, J.A.*, et al.* (2015). Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature *518*, 360-364.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Quivoron, C., Couronne, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.H.*, et al.* (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer cell *20*, 25-38.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proceedings of the National Academy of Sciences of the United States of America *97*, 5237-5242.

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J.*, et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.

Saito, Y., Kitamura, H., Hijikata, A., Tomizawa-Murasawa, M., Tanaka, S., Takagi, S., Uchida, N., Suzuki, N., Sone, A., Najima, Y.*, et al.* (2010). Identification of therapeutic targets for quiescent, chemotherapy-resistant human leukemia stem cells. Science translational medicine *2*, 17ra19.

Samudio, I., Harmancey, R., Fiegl, M., Kantarjian, H., Konopleva, M., Korchin, B., Kaluarachchi, K., Bornmann, W., Duvvuri, S., Taegtmeyer, H.*, et al.* (2010). Pharmacologic inhibition of fatty acid oxidation sensitizes human leukemia cells to apoptosis induction. J Clin Invest *120*, 142-156.

Sarry, J.E., Murphy, K., Perry, R., Sanchez, P.V., Secreto, A., Keefer, C., Swider, C.R., Strzelecki, A.C., Cavelier, C., Recher, C.*, et al.* (2011). Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rgammac-deficient mice. J Clin Invest *121*, 384-395.

Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H.*, et al.* (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature.

Shih, A.H., Abdel-Wahab, O., Patel, J.P., and Levine, R.L. (2012). The role of mutations in epigenetic regulators in myeloid malignancies. Nature reviews Cancer *12*, 599-612.

Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D., Schuh, A.C., Yee, K.W.*, et al.* (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature *506*, 328-333.

Slovak, M.L., Kopecky, K.J., Cassileth, P.A., Harrington, D.H., Theil, K.S., Mohamed, A., Paietta, E., Willman, C.L., Head, D.R., Rowe, J.M.*, et al.* (2000). Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. Blood *96*, 4075-4083.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature reviews Genetics *14*, 204-220.

Tadokoro, Y., Ema, H., Okano, M., Li, E., and Nakauchi, H. (2007). De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. The Journal of experimental medicine *204*, 715-722.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Takeda, A., Goolsby, C., and Yaseen, N.R. (2006). NUP98-HOXA9 induces long-term proliferation and blocks differentiation of primary human CD34+ hematopoietic cells. Cancer Res *66*, 6628-6637.

Taussig, D.C., Miraki-Moud, F., Anjos-Afonso, F., Pearce, D.J., Allen, K., Ridler, C., Lillington, D., Oakervee, H., Cavenagh, J., Agrawal, S.G.*, et al.* (2008). Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. Blood *112*, 568-575.

Taussig, D.C., Vargaftig, J., Miraki-Moud, F., Griessinger, E., Sharrock, K., Luke, T., Lillington, D., Oakervee, H., Cavenagh, J., Agrawal, S.G.*, et al.* (2010). Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. Blood *115*, 1976-1984.

Thol, F., Bollin, R., Gehlhaar, M., Walter, C., Dugas, M., Suchanek, K.J., Kirchner, A., Huang, L., Chaturvedi, A., Wichmann, M.*, et al.* (2014). Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. Blood *123*, 914-920.

Thorsteinsdottir, U., Mamo, A., Kroon, E., Jerome, L., Bijl, J., Lawrence, H.J., Humphries, K., and Sauvageau, G. (2002). Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. Blood *99*, 121-129.

Till, J.E., and Mc, C.E. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiation research *14*, 213-222.

Timp, W., Bravo, H.C., McDonald, O.G., Goggins, M., Umbricht, C., Zeiger, M., Feinberg, A.P., and Irizarry, R.A. (2014). Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. Genome medicine *6*, 61.

Timp, W., and Feinberg, A.P. (2013). Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. Nature reviews Cancer *13*, 497-510.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell stem cell *5*, 442-449.

Valent, P., and Zuber, J. (2014). BRD4: a BET(ter) target for the treatment of AML? Cell cycle *13*, 689-690.

Valk, P.J., Verhaak, R.G., Beijen, M.A., Erpelinck, C.A., Barjesteh van Waalwijk van Doorn-Khosrovani, S., Boer, J.M., Beverloo, H.B., Moorhouse, M.J., van der Spek, P.J., Lowenberg, B.*, et al.* (2004). Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med *350*, 1617-1628.

van Rhenen, A., van Dongen, G.A., Kelder, A., Rombouts, E.J., Feller, N., Moshaver, B., Stigter-van Walsum, M., Zweegman, S., Ossenkoppele, G.J., and Jan Schuurhuis, G. (2007). The novel AML stem cell associated antigen CLL-1 aids in discrimination between normal and leukemic stem cells. Blood *110*, 2659-2666.

Vandiver, A.R., Irizarry, R.A., Hansen, K.D., Garza, L.A., Runarsson, A., Li, X., Chien, A.L., Wang, T.S., Leung, S.G., Kang, S.*, et al.* (2015). Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. Genome biology *16*, 80.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E.*, et al.* (2013). Dynamic DNA methylation across diverse human cell lines and tissues. Genome research *23*, 555-567.

Vassiliou, G.S., Cooper, J.L., Rad, R., Li, J., Rice, S., Uren, A., Rad, L., Ellis, P., Andrews, R., Banerjee, R.*, et al.* (2011). Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. Nat Genet *43*, 470-475.

Waddington, C.H. (2012). The epigenotype. 1942. International journal of epidemiology *41*, 10-13.

Ward, P.S., Patel, J., Wise, D.R., Abdel-Wahab, O., Bennett, B.D., Coller, H.A., Cross, J.R., Fantin, V.R., Hedvat, C.V., Perl, A.E.*, et al.* (2010). The common feature of

leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. Cancer cell *17*, 225-234.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.

Wilson, C.S., Davidson, G.S., Martin, S.B., Andries, E., Potter, J., Harvey, R., Ar, K., Xu, Y., Kopecky, K.J., Ankerst, D.P.*, et al.* (2006). Gene expression profiling of adult acute myeloid leukemia identifies novel biologic clusters for risk classification and outcome prediction. Blood *108*, 685-696.

Wouters, B.J., Lowenberg, B., Erpelinck-Verschueren, C.A., van Putten, W.L., Valk, P.J., and Delwel, R. (2009). Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. Blood *113*, 3088-3091.

Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A.*, et al.* (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nature medicine *20*, 1472-1478.

Yan, X.J., Xu, J., Gu, Z.H., Pan, C.M., Lu, G., Shen, Y., Shi, J.Y., Zhu, Y.M., Tang, L., Zhang, X.W.*, et al.* (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nature genetics *43*, 309-315.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer cell *26*, 577-590.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends in genetics : TIG *13*, 335-340.

Zuber, J., Shi, J., Wang, E., Rappaport, A.R., Herrmann, H., Sison, E.A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M.*, et al.* (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature *478*, 524-528.

Curriculum Vitae

Namyoung Jung

Johns Hopkins School of Medicine

570 Rangos Building, Center for Epigenetics Phone: 443-683-2135

855 North Wolfe Street, Baltimore, MD 21205 E-mail: jnamyoung@gmail.com

---

EDUCATION

2009-present **Johns Hopkins University School of Medicine,** Baltimore, MD

PhD Candidate in Cellular and Molecular Medicine Graduate Program.

Thesis advisor: Andrew P. Feinberg, MD.,MPH.

2003-2007 **Pohang University of Science and Technology (POSTECH)**, Pohang, South Korea

B.S. in Life Science, **Summa Cum Laude**.

Thesis advisor: Yun-je Cho, PhD.

2005 **University of California, Berkeley**, Berkeley, CA

Exchange student in Dept. of Molecular and Cellular Biology.

RESEARCH EXPERIENCE

2009-present **Johns Hopkins University School of Medicine**, Center for Epigenetics and Department of Medicine

PhD candidate, Thesis advisor: Andrew P. Feinberg, MD.,MPH.

Investigated the role of epigenetics, particularly DNA methylation in progenitor cell biology. Two major model systems were explored; Leukemia stem cells (LSCs) along with normal hematopoietic stem and progenitor cells (HSPCs), and induced pluripotent stem cells (iPSCs).

a. Epigenetic features of LSCs and HSPCs using Illumina Infinium HumanMethylation450 bead chip array.

Collaboration with Ravindra Majeti, MD., PhD. at Stanford University School of Medicine.

-Identified LSC epigenetic signature, DNA methylation differences that correlates with differential gene expression between LSCs and blasts.

-Demonstrated prognostic power of LSC epigenetic signature for AML patients.

-Discovered novel regulators for human hematopoiesis with genome-wide DNA methylation and gene expression analysis.

-Identified subgroups of LSCs that reflect cell of origin with DNA methylation analysis.

b. Epigenetic differences of iPSC clones derived by different non-integrating methods using mRNA, sendai virus and episome with whole genome bisulfite sequencing (WGBS) and chromatin immunoprecipitation sequencing (ChIP-seq) analysis.

Collaboration with George Q. Daley, MD., PhD. at Harvard University School of Medicine.

-Identified no significant DNA methylation differences among the iPSCs derived from different methods

2008-2009 **Yonsei University**, Center for Genome Regulation (Seoul, South Korea).

Research assistant, Supervisor: Youngjoon Kim, PhD.

-Optimized native ChIP and identified specific methylation patterns in obesity and cancer.

2007 **Swiss Institute for Experimental Cancer Research (ISREC)**, Telomerase and Chromosome End Replication Laboratory (Lausanne, Switzerland).

Summer research program student. Research advisor: Joachim Lingner, PhD.

-Identified the function of *Rat1,* an exonuclease involved in DNA transcription termination, replication and telomere senescence in *Saccharomyces. Pombe.*

2006-2007 **Pohang University of Science and Technology (POSTECH)** (Pohang, South Korea)

Undergraduate student, Thesis advisor: Yunje Cho, PhD.

-Performed structural study of *MCM10*, a member of DNA replication complex in eukaryote.

2006 **Pohang University of Science and Technology (POSTECH)** (Pohang, South Korea)

Undergraduate student, Research advisor: Sungho Ryu, PhD.

-Studied the relationship between glycolysis and mTOR pathway.

2005 **Pohang University of Science and Technology (POSTECH)** (Pohang, South Korea)

Undergraduate student, Research advisor: Byung-ha Oh, PhD.

-Conducted preliminary structural study of autophagy complex.

HONORS & AWARDS

2014 Mogam Scholarship

2009-2014 Samsung Scholarship

2007 Academic Scholarship for EPFL (Ecole Polytechnique Federale de Lausanne) Summer Research Program

2007 Summa Cum Laude, POSTECH

2007 Nuri Scholarship for excellent research project, KRF (Korea Research Foundation)

2003-2007 Full Scholarship for talented undergraduate students in science major, KOSEF (Korea Science and Engineering Foundation)

TEACHING EXPERIENCE

2009-present **Johns Hopkins University School of Medicine**, Mentor for rotation students.

-Taught how to perform experiments such as bisulfite pyrosequencing, WGBS, comprehensive high-throughput arrays for relative methylation (CHARM), genomic DNA extraction.

-Guided rotation students to make an academic poster and give a talk.

TALKS

Jung N*, Dai B*, Murakami P, Gentles AJ, Majeti R†, Feinberg AP†. Epigenetic signature of leukemia stem cells defines subgroups associated with clinical outcome in AML. The Centers of Excellence in Genomic Science (CEGS) annual meeting, 2014

Jung N*, Dai B*, Murakami P, Irizarry R, Majeti R†, Feinberg AP†. Comprehensive methylome map of human hematopoietic stem and progenitor cells. NHLBI Progenitor Cell Biology Consortium (PCBC) annual meeting, 2014

Jung N*, Dai B*, Murakami P, Irizarry R, Majeti R†, Feinberg AP†. Comprehensive methylome map of human hematopoietic stem and progenitor cells. NHLBI Progenitor Cell Biology Consortium (PCBC) annual meeting, 2013

Jung N*, Dai B*, Murakami P, Irizarry R, Majeti R†, Feinberg AP†. Comprehensive methylome map of human hematopoietic stem and progenitor cells. NHLBI Progenitor Cell Biology Consortium (PCBC) annual meeting, 2012

POSTER PRESENTATIONS

Jung N*, Dai B*, Gentles AJ, Murakami P, Majeti R†, Feinberg AP†. Epigenetic signature of leukemia stem cells defines subgroups associated with clinical outcome and cell of origin in AML. The American Society of Hematology annual meeting, 2014

Jung N*, Dai B*, Murakami P, Irizarry R, Majeti R†, Feinberg AP†. Comprehensive methylome map of human hematopoietic stem and progenitor cells. NHLBI Progenitor Cell Biology Consortium (PCBC) annual meeting, 2014

Jung N*, Dai B*, Murakami P, Irizarry R, Majeti R†, Feinberg AP†. Increased genome-wide epigenetic variation distinguishes acute myeloid leukemia from human normal hematopoiesis, The Centers of Excellence in Genomic Science (CEGS) annual meeting, 2013

Jung N*, Dai B*, Murakami P, Irizarry R, Majeti R†, Feinberg AP†. Comprehensive methylome map of human hematopoietic stem and progenitor cells. NHLBI Progenitor Cell Biology Consortium (PCBC) annual meeting, 2013

Jung N*, Dai B*, Murakami P, Irizarry R, Majeti R†, Feinberg AP†. Increased genome-wide epigenetic variation distinguishes acute myeloid leukemia from human normal hematopoiesis, The Centers of Excellence in Genomic Science (CEGS) annual meeting, 2012

Jung N*, Dai B*, Murakami P, Irizarry R, Majeti R†, Feinberg AP†. Increased genome-wide epigenetic variation distinguishes acute myeloid leukemia from human normal hematopoiesis, Wellcome Trust Epigenomics of Common Diseases Conference, 2012

Jung N*, Dai B*, Murakami P, Irizarry R, Majeti R†, Feinberg AP†. Comprehensive methylome map of human hematopoietic stem and progenitor cells. NHLBI Progenitor Cell Biology Consortium (PCBC) annual meeting, 2012

Jung N*, Dai B*, Ji H, Weissman I, Majeti R†, Feinberg AP†. Epigenetic change during acute myeloid leukemia progression. The Centers of Excellence in Genomic Science (CEGS) annual meeting, 2011

PUBLICATIONS

Jung N*, Li X*, Honaker C, Siegel PB, Andersson L, Feinberg AP. Widespread mutational change driving DNA methylation in size selection of domestic chickens. (In preparation)

Jung N*, Dai B*, Gentles AJ, Majeti R†, Feinberg AP†. A functional DNA methylation signature of acute myeloid leukemia stem cells. (Under Review in *Nature Communication*)

Schlaeger TM, Daheron L, Brickler TR, Entwisle S, Chan K, Cianci A, DeVine A, Ettenger A, Fitzgerald K, Godfrey M, Gupta D, McPherson J, Malwadkar P, Gupta M, Bell B, Doi A, **Jung N**, Li X, Lynes MS, Brookes E, Cherry AB, Demirbas D, Tsankov AM, Zon LI, Rubin LL, Feinberg AP, Meissner A, Cowan CA, Daley GQ. A comparison of non-integrating reprogramming methods. *Nature Biotechnology*. (2014)

Sinha S, Thomas D, Yu L, Gentles A, **Jung N**, Corces-Zimmerman MR, Chan SM, Reinisch A, Feinberg AP, Dill DL, Majeti R. Mutant WT1 is associated with DNA hypermethylation of PRC2 targets in AML and responds to EZH2 inhibition. *Blood*. (2014)

Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LI, Yabuuchi A, Takeuchi A, Cunniff KC, Hongguang H, McKinney-Freeman S,Naveiras O, Yoon TJ, Irizarry RA, **Jung N**, Seita J, Hanna J, Murakami P, Jaenisch R, Weissleder R, Orkin SH, Weissman IL, Feinberg AP, Daley GQ. Epigenetic memory in induced pluripotent stem cells. *Nature* 467:285-90. (2010)

Jung NY, Bae WJ, Chang JH, Kim YC, Cho Y, Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of the central Zn-binding domain of the human MCM10 DNA replication factor, *Acta Crystallogr Sect F Struct Biol Cryst Commun.*(2008)